

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-143912

(43) 公開日 平成11年(1999) 5月28日

(51) Int.Cl.⁹

G 0 6 F 17/30
17/21

識別記号

F I

G 0 6 F 15/40
15/20
15/401

3 7 0 A
5 7 0 R
5 7 0 D
3 1 0 D
3 1 0 A

審査請求 未請求 請求項の数22 O L (全 41 頁) 最終頁に続く

(21) 出願番号 特願平10-82270

(22) 出願日 平成10年(1998) 3月27日

(31) 優先権主張番号 特願平9-242247

(32) 優先日 平9(1997) 9月8日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(72) 発明者 津田 宏

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(72) 発明者 内野 寛治

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(72) 発明者 松井 くにお

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(74) 代理人 弁理士 大菅 義之 (外1名)

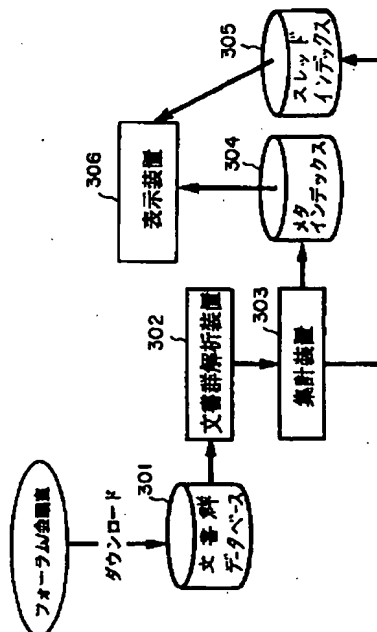
(54) 【発明の名称】 関連文書表示装置

(57) 【要約】

【課題】 大量の文書が含まれる文書集合からの、必要な文書又は必要な文書が含まれているであろう文書群の検索を、支援することにある。

【解決手段】 フォーラム/会議室内の文書群は、所定のサーバコンピュータ上の文書群データベース301として、蓄積される。文書群解析装置302は、文書群データベース301内の各会議室に対応する文書集合毎に、それに含まれる文書群の解析を行う。集計装置303は、文書群解析装置302による解析結果に基づいて、メタインデックス304、スレッドインデックス305、及び索引ファイル404を生成する。表示装置306は、メタインデックス304とスレッドインデックス305を用いて、キーワードビュー、スレッドビュー、又は発言者ビューの何れかの表示形態で、文書群を表示する。

本発明の実施の形態のシステム構成図
(その1)



【特許請求の範囲】

【請求項1】 参照関係を有する文書からなる文書群を表示する関連文書表示装置であって、文書集合を構成する複数の文書間の参照関係を解析することによって、前記文書集合を構成する複数の文書を参照関係を有する1つ以上の文書群に分類し、その分類に関する情報を文書群情報として抽出する文書群解析手段と、

前記文書群情報を参照することにより、前記文書群毎に、その文書群を構成する文書に含まれるキーワードを抽出する文書群キーワード抽出手段と、前記文書群毎に、それに関連するタイトルと、該文書群に対して前記文書群キーワード抽出手段によって抽出されたキーワードとを対応付けて表示する文書群キーワード表示手段と、を含むことを特徴とする関連文書表示装置。

【請求項2】 請求項1に記載の装置であって、前記文書群キーワード表示手段は、ユーザが指定した検索指定に該当する文書が含まれる文書群又は該検索指定に該当するキーワードを強調表示する、ことを特徴とする関連文書表示装置。

【請求項3】 参照関係を有する文書からなる文書群を表示する関連文書表示装置であって、文書集合を構成する複数の文書間の参照関係を解析することによって、前記文書集合を構成する複数の文書を参照関係を有する1つ以上の文書群に分類し、その分類に関する情報を文書群情報として抽出する文書群解析手段と、

前記文書集合を構成する複数の文書のそれぞれから、その文書の属性を示す文書属性情報を抽出する文書属性解析手段と、前記文書群情報及び前記文書属性情報を参照することによって、前記文書群毎に、その文書群の参照関係を、その文書群を構成する各文書に対応する各文書属性情報又はその省略された情報が各ノードとして表示されるツリー構造によって表示する文書群構造表示手段と、を含むことを特徴とする関連文書表示装置。

【請求項4】 請求項3に記載の装置であって、前記文書群構造表示手段は、前記文書群情報及び前記文書属性情報を参照することによって、前記文書群毎に、その文書群の参照関係を、その文書群を構成する各文書に対応する各省略された作者名と各文書量情報とが組み合わされて各ノードとして表示されるツリー構造によって表示する、ことを特徴とする関連文書表示装置。

【請求項5】 請求項3に記載の装置であって、前記文書群構造表示手段は、前記文書群毎に、その文書群に含まれる文書から抽出される複数のトピックを、その文書群に対応して表示される前記ツリー構造を構成する前記各ノードと関連付けて、更に表示する、

ことを特徴とする関連文書表示装置。

【請求項6】 請求項5に記載の装置であって、前記文書群構造表示手段は、前記文書群毎に、前記各トピックとそれに関連する前記各ノードとを、同一の色で表示する、

ことを特徴とする関連文書表示装置。

【請求項7】 請求項3に記載の装置であって、前記文書群構造表示手段は、ユーザが指定した検索指定に該当する文書に対応するノードを強調表示する、ことを特徴とする関連文書表示装置。

【請求項8】 参照関係を有する文書からなる文書群を表示する関連文書表示装置であって、文書集合を構成する複数の文書間の参照関係を解析することによって、前記文書集合を構成する複数の文書を参照関係を有する1つ以上の文書群に分類し、その分類に関する情報を文書群情報として抽出する文書群解析手段と、

該分類された各文書群を、それぞれを構成する文書から抽出されるトピックに基づいて更に細分類し、その細分類に関する情報をトピック分類情報として抽出するトピック解析手段と、

前記文書群情報及び前記トピック分類情報を参照することにより、前記文書群毎及びそれを細分類した前記トピック毎に、そのトピックに関連する文書に含まれるキーワードを抽出するトピックキーワード抽出手段と、前記文書群毎及びそれを細分類した前記トピック毎に、それに関連するタイトルと、該トピックに対して前記トピックキーワード抽出手段によって抽出されたキーワードとを対応付けて表示するトピックキーワード表示手段と、

を含むことを特徴とする関連文書表示装置。

【請求項9】 コンピュータにより使用されたときにそれによって読み出されるプログラムを記録した記録媒体であって、

文書集合を構成する複数の文書間の参照関係を解析することによって、前記文書集合を構成する複数の文書を参照関係を有する1つ以上の文書群に分類し、その分類に関する情報を文書群情報として抽出する機能と、

前記文書群情報を参照することにより、前記文書群毎に、その文書群を構成する文書に含まれるキーワードを抽出する機能と、

前記文書群毎に、それに関連するタイトルと、該文書群に対して前記文書群キーワード抽出手段によって抽出されたキーワードとを対応付けて表示する機能と、を前記コンピュータに行わせるためのプログラムを記録したコンピュータ読出し可能記録媒体。

【請求項10】 コンピュータにより使用されたときにそれによって読み出されるプログラムを記録した記録媒体であって、

文書集合を構成する複数の文書間の参照関係を解析する

ことによって、前記文書集合を構成する複数の文書を参照関係を有する1つ以上の文書群に分類し、その分類に関する情報を文書群情報として抽出する機能と、前記文書集合を構成する複数の文書のそれぞれから、その文書の属性を示す文書属性情報を抽出する機能と、前記文書群情報及び前記文書属性情報を参照することによって、前記文書群毎に、その文書群の参照関係を、その文書群を構成する各文書に対応する各文書属性情報又はその省略された情報が各ノードとして表示されるツリー構造によって表示する機能と、を前記コンピュータに行わせるためのプログラムを記録したコンピュータ読出し可能記録媒体。

【請求項11】 参照関係を有する文書からなる文書群を表示する関連文書表示装置であって、参照関係を有する文書群からなる文書データベースから、文書内容及び作者間の投稿パターンに基づいて前記文書群の内容を推定して、該内容の話題パターンに対応するインデックスを生成する内容推定手段と、利用者からの前記文書データベースに対する検索要求を入力する入力手段と、前記文書データベースの文書を検索する検索エンジン手段と、該検索エンジン手段からの検索結果と前記インデックスを利用して1つ以上のビューを生成し、該1つ以上のビューを切り替えて表示装置に表示するビュー生成手段と、を含むことを特徴とする関連文書表示装置。

【請求項12】 請求項11に記載の装置であって、前記ビュー生成手段は、表示される文書において同一の属性を有する文書をまとめることにより、表示される文書間の参照関係の木構造及び該各文書に対応するノードに関する情報を簡略表示することにより、利用者に表示される文書間の参照関係の全体構造を容易に把握させる、ことを特徴とする関連文書表示装置。

【請求項13】 請求項11に記載の装置であって、前記ビュー生成手段は、画面のサイズに合わせて、表示される文書間の参照関係において参照数の多い文書のみを選択することにより、表示される文書間の参照関係の木構造及び該各文書に対応するノードに関する情報を簡略表示することにより、利用者に表示される文書間の参照関係の全体構造を容易に把握させる、ことを特徴とする関連文書表示装置。

【請求項14】 請求項11に記載の装置であって、前記ビュー生成手段は、利用者が入力したキーワードを含む文書であってそれに対する参照数が多い文書のみを選択することにより、表示される文書間の参照関係の木構造及び該各文書に対応するノードに関する情報を簡略表示することにより、利用者に文書データベースに対する検索結果の全体構造を容易に把握させる、

ことを特徴とする関連文書表示装置。

【請求項15】 請求項11に記載の装置であって、前記ビュー生成手段は、表示される文書間の参照関係を、該各文書に対して前記内容推定手段が推定している話題パターンと共に表示することにより、利用者に文書データベースに対する検索結果の全体構造を容易に把握させる、

ことを特徴とする関連文書表示装置。

【請求項16】 請求項11に記載の装置であって、前記ビュー生成手段は、表示される文書のうち特定の時間区間の文書のみをカレンダー形式で表示することにより、利用者に文書データベースに対する検索結果を時間的情報と関連付けて容易に把握させる、

ことを特徴とする関連文書表示装置。

【請求項17】 請求項11に記載の装置であって、前記ビュー生成手段は、前記内容推定手段が推定した特定の話題パターンを強調して表示することにより、利用者に文書データベース内の文書の参照関係の重要部分のみを容易に把握させる、

ことを特徴とする関連文書表示装置。

【請求項18】 請求項11に記載の装置であって、前記ビュー生成手段は、前記検索エンジン手段に対して、前記内容推定手段が推定した特定の話題パターンのうち、質問と回答に相当する文書のみを対象として検索を実行させることにより、利用者の質問事項に対応する質問及び回答の対を容易に把握させる、

ことを特徴とする関連文書表示装置。

【請求項19】 請求項11に記載の装置であって、前記ビュー生成手段は、前記文書データベース内の各文書毎の投稿履歴と前記内容推定手段が推定した特定の話題パターンに基づいて、特定の作者を強調表示して表示することにより、利用者に文書データベース内の文書の参照関係の重要部分のみを容易に把握させる、

ことを特徴とする関連文書表示装置。

【請求項20】 請求項11に記載の装置であって、前記ビュー生成手段は、前記文書データベース内の各文書の作者をノードとし、該文書間の参照関係をリンク、参照回数をリンクの強調度とする有向グラフによって表示することにより、利用者に文書データベース内の文書の参照関係を作者の観点から容易に把握させる、

ことを特徴とする関連文書表示装置。

【請求項21】 請求項11に記載の装置であって、前記文書データベースに蓄積される文書は、ネットワークを通じてダウンロードされるネットワークニュースの記事文書である、

ことを特徴とする関連文書表示装置。

【請求項22】 コンピュータにより使用されたときにそれによって読み出されるプログラムを記録した記録媒体であって、参照関係を有する文書群からなる文書データベースか

ら、文書内容及び作者間の投稿パターンに基づいて前記文書群の内容を推定して、該内容の話題パターンに対応するインデックスを生成する機能と、利用者からの前記文書データベースに対する検索要求を入力する機能と、前記文書データベースの文書を検索する機能と、該検索結果と前記インデックスを利用して1つ以上のビューを生成し、該1つ以上のビューを切り替えて表示装置に表示する機能と、を前記コンピュータに行わせるためのプログラムを記録したコンピュータ読出し可能記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、参照関係にある文書群を整理し、その文書群をいろいろな観点から表示することによって、ユーザによる必要な情報へのアクセスを支援する技術に関する。

【0002】

【従来の技術及び発明が解決しようとする課題】パソコン通信又はコンピュータネットワーク上で運用される電子会議室や電子ニュース等において、順次蓄積される文書集合の中から必要な文書をより迅速かつ簡単に見つけ出したいという要請が、従来からある。

【0003】このような要請に対して、文書集合中の各文書のタイトルを作成日順に並び替え、その結果得られるタイトルリストをユーザに掲示するという従来技術が知られている。

【0004】また、文書集合を互いに参照関係のある文書から構成される文書群に分類し、文書群中の各文書のタイトルをインデントして表示することにより、各文書の参照関係を掲示する従来技術や、文書群中の各文書の番号をツリーで表示することにより各文書の参照関係を掲示する従来技術も知られている。

【0005】更に、文書集合の中から、特定のキーワードを含む文書を全文検索し、その検索結果を羅列的に掲示する従来技術も知られている。しかし、これらの従来技術におけるような限定された情報表示のみでは、以下に示される問題点を解決することができなかった。

1. 雑多な文書集合の中から必要な文書又は必要な文書が含まれているであろう文書群を見つけ出すためには、掲示される文書のタイトルに頼るしかない。タイトルは、必ずしも文書内容を正確に表わしているとは限らないため、正確な検索が困難である。
2. タイトルのインデント表示や文書番号のツリー表示だけでは、文書群全体の構造を把握すること、及び文書群における話題の推移を把握することが困難である。
3. 種々の観点から必要な文書にアクセスすることができない。
4. 検索結果が多数件ある場合に、更に絞り込み検索を実行するか検索結果のリストを1件1件チェックしなけ

れば、必要な文書にアクセスすることができない。

一方、複数の特定の文書からキーワードを抽出し、共通のキーワードを含む各文書に対し自動的に他の文書へのリンクを設定する技術が、知られている。この従来技術は、特許の公知例や研究論文等の特定文書中で互いの文献を相互参照することを可能にすることによって、関連する複数の文書を効率的に読み広げることが可能にする。

【0006】しかし、このような従来技術は、関連する文書の参照を容易にすることを目的としており、電子会議室や電子ニュース等の文書集合からの、必要な文書又は必要な文書が含まれているであろう文書群の検索の支援に、適用することはできなかった。

【0007】本発明の課題は、大量の文書が含まれる文書集合からの、必要な文書又は必要な文書が含まれているであろう文書群の検索を、支援することにある。

【0008】

【課題を解決するための手段】本発明は、参照関係を有する文書からなる文書群を表示する関連文書表示装置を前提とする。

【0009】本発明の第1の態様は、以下の構成を有する。まず、文書群解析手段（構造解析部402）は、文書集合を構成する複数の文書間の参照関係を解析することによって、文書集合を構成する複数の文書を参照関係を有する1つ以上の文書群（スレッド）に分類し、その分類に関する情報を文書群情報（スレッドインデックス305）として抽出する。

【0010】文書群キーワード抽出手段（内容解析部403）は、文書群情報を参照することにより、文書群毎に、その文書群を構成する文書に含まれるキーワードを抽出する。

【0011】文書群キーワード表示手段（表示装置306）は、文書群毎に、それに関連するタイトルと、その文書群に対して文書群キーワード抽出手段によって抽出されたキーワードとを対応付けて表示する。

【0012】上述した本発明の第1の態様の構成により、キーワードを頼りにして、雑多な文書集合の中から必要な文書が含まれているであろう文書群を容易に見つけ出すことが可能となる。

【0013】本発明の第2の態様は、上述の本発明の第1の態様の構成において、文書群キーワード表示手段が、ユーザが指定した検索指定に該当する文書が含まれる文書群又はその検索指定に該当するキーワードを強調表示する構成を有する。

【0014】上述した本発明の第2の態様の構成により、検索指定に該当する文書群を一目で把握することが可能となる。本発明の第3の態様は、以下の構成を有する。

【0015】文書群解析手段（構造解析部402）は、文書集合を構成する複数の文書間の参照関係を解析する

ことによって、文書集合を構成する複数の文書を参照関係を有する1つ以上の文書群に分類し、その分類に関する情報を文書群情報として抽出する。

【0016】文書属性解析手段（書式解析部401）は、文書集合を構成する複数の文書のそれぞれから、その文書の属性を示す文書属性情報（メタインデックス304）を抽出する。

【0017】文書群構造表示手段（表示装置306）は、文書群情報及び文書属性情報を参照することによって、文書群毎に、その文書群の参照関係を、その文書群を構成する各文書に対応する各文書属性情報又はその省略された情報が各ノードとして表示されるツリー構造によって表示する。より具体的には、文書群構造表示手段は、文書群毎に、その文書群の参照関係を、その文書群を構成する各文書に対応する各省略された作者名と各文書量情報（各行数）とが組み合わされて各ノードとして表示されるツリー構造によって表示する。

【0018】上述した本発明の第3の態様の構成により、文書群全体の構造を容易に把握することが可能となる。本発明の第4の態様は、上述の本発明の第3の態様の構成において、文書群構造表示手段が、文書群毎に、その文書群に含まれる文書から抽出される複数のトピック（メイントピック、サブトピック）を、その文書群に対応して表示されるツリー構造を構成する各ノードと関連付けて、更に表示する構成を有する。より具体的には、文書群構造表示手段は、文書群毎に、各トピックとそれに関連する各ノードとを、同一の色で表示する。

【0019】上述した本発明の第4の態様の構成により、文書群全体の話題の推移を容易に把握することが可能となる。本発明の第5の態様は、前述の本発明の第3の態様の構成において、文書群構造表示手段が、ユーザが指定した検索指定に該当する文書に対応するノードを強調表示する構成を有する。

【0020】上述した本発明の第5の態様の構成により、スレッドの構造を把握しつつ、検索を実行することが可能となる。本発明の第6の態様は、以下の構成を有する。

【0021】文書群解析手段（構造解析部402）は、文書集合を構成する複数の文書間の参照関係を解析することによって、文書集合を構成する複数の文書を参照関係を有する1つ以上の文書群に分類し、その分類に関する情報を文書群情報として抽出する。

【0022】トピック解析手段（構造解析部402）は、その分類された各文書群を、それぞれを構成する文書から抽出されるトピックに基づいて更に細分類し、その細分類に関する情報をトピック分類情報（サブトピックインデックス）として抽出する。

【0023】トピックキーワード抽出手段（内容解析部403）は、文書群情報及びトピック分類情報を参照することにより、文書群毎及びそれを細分類したトピック

毎に、そのトピックに関連する文書に含まれるキーワードを抽出する。

【0024】トピックキーワード表示手段（表示装置306）は、文書群毎及びそれを細分類したトピック毎に、それに関連するタイトルと、そのトピックに対してトピックキーワード抽出手段により抽出されたキーワードとを対応付けて表示する。

【0025】上述した本発明の第6の態様の構成により、キーワードによるより精密なトピックの絞込みを行うことが可能となる。なお、本発明は、コンピュータにより使用されたときに、上述の本発明の各構成によって実現される機能と同様の機能をコンピュータに行わせるためのコンピュータ読出し可能記録媒体として構成することもできる。

【0026】

【発明の実施の形態】以下、図面を参照しながら本発明の実施の形態について詳細に説明する。

全体構成

図1は、本発明の実施の形態が対象とする、文書集合及び文書群の例を示す図である。

【0027】この例では、コンピュータネットワーク上に、ユーザが話題別に議論を行うフォーラムと呼ばれる仮想的な公開討論会場が設けられており、各フォーラムは、会議室と呼ばれる、更に細分化された話題を扱う複数の仮想的な会場に分類されている。ユーザがこの会議室に、発言文書を投稿（アップロード）することによって、議論が進行する。フォーラム及び会議室とも、サーバコンピュータ上のストレージエリアとして構成され、それらには、上述した分類基準に従って、文書が蓄積される。また、各会議室において、互いに参照関係を有する複数の文書からなる文書群がスレッドを構成する。

【0028】ユーザが投稿する文書は、例えば、図2に示されるデータ構造を有しており、その文書の番号を示す文書番号、日付、タイトル、その文書が参照する文書の番号である参照文書番号、作者名（発言者名）等の文書の属性フィールドが記載されるヘッダ部と、文書の本体が記載される内容部とから構成されている。

【0029】本発明の実施の形態では、以下のような表示形態が可能となる。

1. キーワードビュー：会議室を構成する文書集合において、その文書集合中の各スレッド毎に、そのスレッドを構成する文書群からキーワードが抽出され、それらのキーワードが、それらの文書数及びそれらが含まれるスレッドのタイトルと合わせて、図25に示される表示形態で表示される。

【0030】キーワードビューによって、ユーザは、キーワードを頼りにして、雑多な文書集合の中から必要な文書が含まれているであろうスレッド（文書群）を容易に見つけ出すことが可能となる。

2. スレッドビュー：文書の参照関係、タイトル、作者

名、及び行数が一目にわかる図26に示される表示形態で、各スレッドを構成する文書群が表示される。

【0031】スレッドビューにより、スレッド全体の構造を把握し話題の推移を容易に把握することが可能となる。

3. 発言者ビュー：各文書のタイトルが、発言者（作者）毎に分類され、かつ発言者が発言の多い順にソートされ、同一発言者内では日付順で、図27に示される表示形態で、表示される。

【0032】発言者及び発言日付という観点から、文書集合（会議室）内の文書を参照することが可能となる。

4. 各ビューへの検索結果の反映：ユーザが指定した検索キーワードに関連する文書が、図32又は図33に示される表示形態で、キーワードビュー、スレッドビュー等の表示中で強調表示される。

【0033】この結果、より正確な文書の把握が可能となる。

5. 各ビューの切替え機能：上述のキーワードビュー、スレッドビュー、及び発言者ビューが任意に切替え可能とされることにより、種々の観点から必要な文書にアクセス可能となる。

以上のような表示形態を可能とする本発明の実施の形態について、詳細に説明する。

【0034】図3及び図4は、本発明の実施の形態のシステム構成図である。フォーラム／会議室内の文書群は、所定のサーバコンピュータ上の文書群データベース301として、蓄積される。

【0035】文書群解析装置302は、文書群データベース301内の各会議室に対応する文書集合毎に、それに含まれる文書群の解析を行う。集計装置303は、文書群解析装置302による解析結果に基づいて、メタインデックス304、スレッドインデックス305、及び索引ファイル404を生成する。

【0036】表示装置306は、メタインデックス304とスレッドインデックス305を用いて、キーワードビュー、スレッドビュー、又は発言者ビューの何れかの表示形態で、文書群を表示する。

【0037】また、文字列検索装置405は、ユーザによる検索語の指定に基づいて、索引ファイル404を用いながら文書群データベース301内の文書集合を構成する各文書に対して検索を実行する。表示装置306は、その検索結果を、キーワードビュー又はスレッドビューに反映させて表示する。

【0038】表示装置306は、メタインデックス304とスレッドインデックス305を用いて、キーワードビュー、スレッドビュー、又は発言者ビューの何れかの表示形態で、文書群を表示する。

【0039】文書群解析装置302は、書式解析部401、構造解析部402、及び内容解析部403とから構成される。書式解析部401は、文書群データベース3

01内の文書集合を構成する図2のデータ構造を有する各文書のヘッダ部から、文書番号、タイトル、作者名、日付、及び参照文書番号を抽出し、また、各文書の内容部の行数を算出し、それらを、集計装置303を経由して、図5に示されるデータ構造を有するメタインデックス304に登録する。

【0040】構造解析部402は、書式解析部401が各文書から抽出した文書番号と参照文書番号に基づいて、各文書をスレッドを単位とする文書群に分類し、集計装置303を経由して、スレッド毎に、それを構成する文書の参照関係のリストであるスレッドインデックス305を作成する。

【0041】図6は、スレッドインデックス305のデータ構造を示す図である。各スレッド毎に、ルート文書番号と、文書数と、スレッドの構造を示すリストとが登録される。リストは、（親文書番号 子文書番号／サブツリー 子文書番号／サブツリー...）という記述形式によって記述され、“子文書番号／サブツリー”の部分には、更に再帰的（リカーシブ）に、子リストを記述することができる。

【0042】図6に例示される2つのスレッドの各リストにより表現される各参照関係は、同図の表の右側に示される如くである。また、構造解析部402は、解析したスレッドを構成する文書群を、更にタイトルを同一とするサブ文書群に分類し、各サブ文書群に色番号を付与し、各サブ文書群に含まれる文書に対応する図5のデータ構造を有するメタインデックス304内のエントリに、その文書が属するサブ文書群に付与された色番号を登録する。

【0043】内容解析部403は、構造解析部402によって分類されたスレッド毎に、そのスレッドを構成する文書群を1つの結合文書ファイルにまとめ、その結合文書からキーワードを抽出する。日本語文書からキーワードを抽出する技術としては、種々の公知技術を採用することができる。この場合に、キーワード抽出の精度を向上させるために、ノイズとなる文字がパターンマッチングによって除去される。また、例えば、上位所定個数のキーワードのみが抽出される。

【0044】内容解析部403によって抽出された各スレッドのキーワードは、集計装置303を経由して、そのスレッドのルート文書に対応する図5のデータ構造を有するメタインデックス304のエントリに、登録される。

【0045】また、内容解析部403は、スレッド毎に抽出したキーワードから、そのキーワードに含まれる索引語を抽出し、集計装置303を経由して、図7に示されるデータ構造を有する索引ファイル404を生成する。

【0046】この索引ファイル404は、前述したように、文字列検索装置405によって参照される。

文書群解析装置302の詳細説明

図8は、図4の文書群解析装置302内の書式解析部401及び構造解析部402が実現する制御動作を示す動作フローチャートである。

【0047】まず、書式解析部401は、文書群データベース301から新たに登録された新規文書ファイルから文書データを1行ずつ読み込みながら、その文書ファイルのヘッダ部（図2参照）から、文書番号、タイトル、作者名、日付、及び参照文書番号を抽出する（ステップ801→802→803→801のループ）。

【0048】書式解析部401は、ヘッダ部の抽出を終了すると、集計装置303を経由して、図5のデータ構造を有するメタインデックス304において新規エントリを生成し、そのエントリに、抽出した文書番号、タイトル、作者名、日付、及び参照文書番号を登録する（ステップ802→804）。

【0049】次に、書式解析部401は、上記新規文書ファイル内のヘッダ部以降の内容部（図2参照）から文書データを1行ずつ読み込みながら、文書末尾（E.O.F.: エンドオブファイル）が検出されるまで、内容部の行数をカウントする（ステップ805→806→807→805のループ）。

【0050】書式解析部401は、文書末尾を検出すると、集計装置303を経由して、それまでにカウントした内容部の行数を、図5のデータ構造を有するメタインデックス304の、現在処理中の新規文書の文書番号に対応するエントリに、登録する（ステップ806→808）。

【0051】なお、上述の行数のカウント処理において、他の文書から引用している行（例えば”>”で始まる行）については、行数のカウントには算入しないことによって、その文書が実質的に発言している行数をカウントするように構成されてもよい。

【0052】続いて、構造解析部402に制御が移り、構造解析部402は、まず、現在処理中の新規文書の文書番号を、集計装置303を経由して、図6のデータ構造を有するスレッドインデックス305に登録する（ステップ809）。

【0053】図9は、上記ステップ809の登録動作を示す動作フローチャートである。まず、構造解析部402は、現在処理中の新規文書が、或るスレッドのルート文書であるかを判定する（ステップ901）。具体的には、構造解析部402は、図8のステップ801～803のループにおいて、現在処理中の新規文書から参照文書番号が検出されなかった場合に、その文書はルート文書であると判定する。

【0054】構造解析部402は、現在処理中の新規文書が或るスレッドのルート文書であると判定した場合には、集計装置303を経由して、スレッドインデックス305において新規エントリを生成し、そのエントリに

現在処理中の新規文書の文書番号をルート文書番号として登録する（ステップ901→902）。

【0055】構造解析部402は、ステップ902の処理の後、上記エントリの文書数を1に初期設定し（ステップ906）、図8のステップ809の処理を終了する。一方、構造解析部402は、現在処理中の新規文書が或るスレッドのルート文書ではないと判定した場合には、現在処理中の新規文書のヘッダ部から抽出されている参照文書番号が含まれるスレッドインデックス305中のエントリに、その参照文書番号を親文書番号とするリストが存在するかを判定する（ステップ901→903）。

【0056】構造解析部402は、上述のエントリに、現在処理中の新規文書のヘッダ部から抽出されている参照文書番号を親文書番号とするリストが存在すると判定した場合には、現在処理中の新規文書のヘッダ部から抽出されている文書番号を、そのリストの子文書番号として登録する（ステップ903→905）。

【0057】一方、構造解析部402は、上述のエントリに、現在処理中の新規文書のヘッダ部から抽出されている参照文書番号を親文書番号とするリストが存在しないと判定した場合には、そのエントリに、その参照文書番号を親文書番号とするリストを生成した上で、現在処理中の新規文書のヘッダ部から抽出されている文書番号を、そのリストの子文書番号として登録する（ステップ903→904→905）。

【0058】構造解析部402は、ステップ905の処理の後、上述のエントリの文書数を更新（プラス1）し（ステップ906）、図8のステップ809の処理を終了する。

【0059】上記図9の動作フローチャートによって実現される制御動作の具体例につき、図10の説明図を用いて説明する。この図は、図6のスレッドインデックス305において、ルート文書番号が”001”であるスレッドのエントリのリストが生成される過程を示すものである。

【0060】まず、文書番号”001”のルート文書が処理される時点で、図9のステップ901→902が実行されることにより、スレッドインデックス305において新規エントリが生成され、そのエントリに文書番号”001”がルート文書番号として登録され（図10の（1））、上記エントリの文書数が1に初期設定される。

【0061】次に、文書番号”002”の文書が処理される時点で、図9のステップ901→903→904→905→906が実行されることにより、スレッドインデックス305内のルート文書番号”001”のエントリにおいて、文書番号”002”の文書から抽出された参照文書番号”001”を親文書番号とするリストが生成された後（図10の（2））、文書番号”002”が

そのリストの子文書番号として登録され（図10の（3）の下線部）、上記エントリの文書数が2に更新される。

【0062】次に、文書番号”003”の文書が処理される時点で、図9のステップ901→903→904→905→906が実行されることにより、スレッドインデックス305内のルート文書番号”001”のエントリにおいて、文書番号”003”の文書から抽出された参照文書番号”002”を親文書番号とするリストが生成された後（図10の（4）の下線部）、文書番号”003”がそのリストの子文書番号として登録され（図10の（5）の下線部）、上記エントリの文書数が3に更新される。

【0063】次に、文書番号”004”の文書が処理される時点で、図9のステップ901→903→905→906が実行されることにより、スレッドインデックス305内のルート文書番号”001”のエントリにおいて、文書番号”004”の文書から抽出された参照文書番号”001”を親文書番号とするリストの子文書番号として、文書番号”004”が登録され（図10の

（6）の下線部）、上記エントリの文書数が4に更新される。

【0064】次に、文書番号”005”の文書が処理される時点で、図9のステップ901→903→904→905→906が実行されることにより、スレッドインデックス305内のルート文書番号”001”のエントリにおいて、文書番号”005”の文書から抽出された参照文書番号”004”を親文書番号とするリストが生成された後（図10の（7）の下線部）、文書番号”005”がそのリストの子文書番号として登録され（図10の（8）の下線部）、上記エントリの文書数が5に更新される。

【0065】最後に、文書番号”006”の文書が処理される時点で、図9のステップ901→903→904→905→906が実行されることにより、スレッドインデックス305内のルート文書番号”001”のエントリにおいて、文書番号”006”の文書から抽出された参照文書番号”005”を親文書番号とするリストが生成された後（図10の（9）の下線部）、文書番号”006”がそのリストの子文書番号として登録され（図10の（10）の下線部）、上記エントリの文書数が6に更新される。

【0066】以上説明した図8のステップ809の処理の後、構造解析部402は、現在処理中の新規文書について色番号を決定し、その色番号を図5のデータ構造を有するメタインデックス304中の上記新規文書の文書番号に対応するエントリに登録する処理を実行する（図8のステップ810）。

【0067】図11は上記ステップ810の登録動作を示す動作フローチャートである。なお、この登録動作で

は、図12に示されるデータ構造を有するカラーテーブルが使用される。このテーブルは、特に図示しない記憶装置に記憶される。

【0068】まず、構造解析部402は、現在処理中の新規文書が、或るスレッドのルート文書であるか否かを判定する（ステップ1101）。具体的には、構造解析部402は、図8のステップ801～803のループにおいて、現在処理中の新規文書から参照文書番号が検出されなかった場合に、その文書はルート文書であると判定する。

【0069】構造解析部402は、現在処理中の新規文書が或るスレッドのルート文書であると判定した場合は、そのルート文書の文書番号に対応するエントリを、図12のデータ構造を有するカラーテーブルに登録し、そのエントリに現在処理中の新規文書から抽出された文書番号及びタイトル（図8のステップ804参照）と、初期色番号を登録する（ステップ1101→1102→1103）。図12の例では、ルート文書番号”001”の色番号”#1”に対応するエントリが登録され、そのタイトルはメインピックとなり、また、そのエントリの文書番号フィールドには、当初はルート文書番号”001”のみが登録される。

【0070】その後、構造解析部402は、図5のデータ構造を有するメタインデックス304中の上記新規文書の文書番号に対応するエントリに、ステップ1103で登録した初期色番号を登録し（ステップ1103→1110）、図8のステップ810の処理を終了する。

【0071】一方、構造解析部402は、現在処理中の新規文書が或るスレッドのルート文書ではないと判定した場合には、現在処理中の新規文書から抽出されたタイトル（図8のステップ804参照）が、”Re:”等の参照記号を含んでいるか否かを判定する（ステップ1101→1104）。

【0072】構造解析部402は、現在処理中の新規文書から抽出されたタイトルが参照記号を含んでいると判定した場合はそのタイトルから参照記号を削除し（ステップ1104→1105）、現在処理中の新規文書から抽出されたタイトルが参照記号を含んでいないと判定した場合にはステップ1105は実行しない。

【0073】その後、構造解析部402は、図12のデータ構造を有するカラーテーブル中の現在処理中の新規文書が属するスレッドに対応する何れかのエントリに、現在処理中の新規文書から抽出され参照記号を含まないタイトルと同じタイトルが登録されているか否かを判定する（ステップ1106）。現在処理中の新規文書が属するスレッドとそのルート文書番号は、図8のステップ809の処理において図6のデータ構造を有するスレッドインデックス305のエントリが決定される際に検出されるため、そのルート文書番号からカラーテーブル中のエントリが決定される。例えば、現在処理中の新規文

書が文書番号”002”の文書である場合には、図12に示されるカラーテーブルにおいて、ルート文書番号”001”に属するエントリが検出される。

【0074】構造解析部402は、図12のデータ構造を有するカラーテーブル中の現在処理中の新規文書が属するスレッドに対応する何れかのエントリに、現在処理中の新規文書から抽出され参照記号を含まないタイトルと同じタイトルが登録されていると判定した場合には、そのエントリの文書番号フィールドに、現在処理中の新規文書の文書番号を登録する（ステップ1106→1107）。例えば、現在処理中の新規文書が文書番号”002”の文書である場合には、図12に示されるカラーテーブルにおいて、ルート文書番号”001”に属し色番号”#1”が登録されているエントリの文書番号フィールドに、文書番号”002”が登録される。

【0075】その後、構造解析部402は、図5のデータ構造を有するメタインデックス304中の上記新規文書の文書番号に対応するエントリに、ステップ1107で登録が行われたカラーテーブル中のエントリに設定されている色番号を登録し（ステップ1107→1110）、図8のステップ810の処理を終了する。

【0076】一方、構造解析部402は、図12のデータ構造を有するカラーテーブル中の現在処理中の新規文書が属するスレッドに対応する何れのエントリにも、現在処理中の新規文書から抽出され参照記号を含まないタイトルと同じタイトルが登録されていないと判定した場合は、カラーテーブルにおいて上記スレッドに対応する新たなエントリを作成し（ステップ1108）、その作成したエントリに、そのスレッド内で新たな色番号と、現在処理中の新規文書から抽出された文書番号及びタイトル（図8のステップ804参照）を登録する（ステップ1106→1108→1109）。例えば、現在処理中の新規文書が文書番号”003”の文書である場合には、図12のカラーテーブルにおいて、ルート文書番号”001”に属する新たなエントリが作成され、そのエントリに、色番号”#2”と、文書番号”003”の文書のタイトルと、文書番号”003”とが登録される。このタイトルは、ルート文書番号”001”のタイトルであるメインピックに対して、サブピック1となる。

【0077】その後、構造解析部402は、図5のデータ構造を有するメタインデックス304中の上記新規文書の文書番号に対応するエントリに、ステップ1109でカラーテーブル中の新たなエントリに設定された新たな色番号を登録し（ステップ1109→1110）、図8のステップ810の処理を終了する。

【0078】内容解析部403は、図6のデータ構造を有するスレッドインデックス305を参照することにより、前述したように、スレッド毎に、そのスレッドを構成する文書群を1つの結合文書ファイルにまとめ、その

結合文書からキーワードを抽出する。この結果、抽出された各スレッドのキーワードは、そのスレッドのルート文書に対応する図5のデータ構造を有するメタインデックス304のエントリに、登録される。

表示装置306の詳細説明

表示装置306は、前述したように、図5のデータ構造を有するメタインデックス304と図6のデータ構造を有するスレッドインデックス305を用いて、キーワードビュー、スレッドビュー、又は発言者ビューの何れかの表示形態で、文書群を表示することができる。

【0079】ここで例えば、図4のシステムが、ホームページの表示を制御するWebサーバに接続されるように構成されれば、ユーザは、パーソナルコンピュータ等の手元の端末上のWebブラウザアプリケーションから上記Webサーバに接続して特定のフォーラムの特定の会議室にログインした後に、所定の各GUI（グラフィックユーザインタフェース）ボタンをマウス装置等でクリックすることによって、キーワードビュー、スレッドビュー、又は発言者ビューを切り替えて表示させることができる。

【0080】より具体的には、表示装置306は、Webサーバに対して例えばCGI（コモンゲートウェイインタフェース）アプリケーションとして機能し、Webサーバから引き渡されたユーザからのリクエストに 응답して、キーワードビュー、スレッドビュー、又は発言者ビュー等の各ビューを表現するHTML（ハイパーテキストマークアップ言語）による文書データを生成し、それをWebサーバに引き渡す。そして、これらのHTML文書データをWebサーバがユーザにインターネット等のコンピュータネットワークを経由して返信することにより、ユーザの端末上のWebブラウザアプリケーションに、上記ビューが表示される。

【0081】まず、表示装置306が実現するキーワードビューの表示動作について説明する。前述したようにキーワードビューにおいては、スレッド毎に、そのスレッドを構成する文書群から抽出されているキーワードが、その文書群の文書数及びそのスレッドのタイトルと合わせて、図25に示される表示形態で表示される。

【0082】図13は、表示装置306が実行するキーワードビューの表示動作を示す動作フローチャートである。。まず、表示装置306は、図5のデータ構造を有するメタインデックス304のファイルを読み込む（ステップ1301）。

【0083】次に、表示装置306は、メタインデックス304のファイルから1エントリずつデータを読み込みながら、ルート文書が登録されているエントリを検索する（ステップ1301→1302→1301のループ）。各エントリがルート文書が登録されているエントリであるか否かは、各エントリの参照文書番号フィールドの値が無効なデータ値であるか否かによって判定する

ことができる。

【0084】表示装置306は、ルート文書が登録されているエントリを検出すると、そのルート文書番号を、そのルート文書番号に対応する文書群データベース301内のルート文書を表示するためのアプリケーションへの統一されたアドレス情報であるURL (Uniform Resource Locator) がHREF属性の値として指定されるアンカータグに変換する(ステップ1302→1303)。

【0085】次に表示装置306は、図6のデータ構造を有するスレッドインデックス305において、上記ルート文書番号に対応するエントリを参照することにより、そのスレッドに含まれる文書数(子文書数)を取得する(ステップ1304)。

【0086】そして、表示装置306は、図5のデータ構造を有するメタインデックス304において、上記ルート文書が登録されているエントリから、タイトル(メイントピック)と、キーワードとを抽出し、それらと、ステップ1303で変換されたアンカータグ形式のルート文書番号、及びステップ1304で取得した子文書数からなるデータ列を1テーブルレコードとして含むHTMLテーブル文書データを作成する(ステップ1305)。

【0087】続いて、表示装置306は、メタインデックス304のファイルから文書末尾(EOF)を検出するまで、上記ステップ1301~1305の一連の処理を繰り返し実行することにより、各スレッド毎のHTMLテーブル文書データを作成する(ステップ1306→1301)。

【0088】表示装置306は、メタインデックス304のファイルから文書末尾を検出すると(ステップ1306の判定がYES)、最終的に得られたHTMLテーブル文書データをWebサーバに引き渡して、キーワードビューの表示動作を終了する。この結果、ユーザの端末のWebブラウザアプリケーション上に、図25に例示されるようなテーブル形式で、キーワードビューが表示される。

【0089】ユーザは、キーワードビュー上の各スレッド毎のキーワードを頼りにして、雑多な文書集合の中から必要な文書が含まれているであろうスレッドを容易に見つけ出すことが可能となる。

【0090】また、ユーザは、ルート文書に対応するアンカーをマウス装置等でクリックすることによって、所望のスレッドのルート文書に即座にアクセスすることができる。

【0091】上述のキーワードビューの表示動作において、子文書数に応じて、各スレッドのテーブルレコードを色分けして表示するように構成されてもよい。これによって、ユーザは、スレッド毎の発言数を一目で判別することができる。

【0092】続いて、表示装置306が実現するスレッドビューの表示動作について説明する。前述したように、スレッドビューにおいては、文書の参照関係、タイトル、作者名、及び行数が一目にわかる図26に示される表示形態で、各スレッドを構成する文書群が表示される。

【0093】図26において、スレッドの参照関係及び話題の推移が色付きツリーによって表示される。各ツリーのノードは、各文書に対応し、その文書の作者名の先頭文字(2バイト)とその文書の行数を用いて、作者名[行数]の形式で表示される。また、各ノードの前後には、“*”、“+”、“=”、又は“.”等の記号が付される。これらの記号の意味は、下記の通りである。

“*” この記号が付される文書がルート文書である。

“+” この記号が付される文書が参照している文書が他の文書によっても参照されている。

“=” この記号が付される文書を参照している文書が存在する。

“.” この記号が付される文書を参照している文書が存在しない。また、図26において、“Main Topic:”に続いてそのスレッドのルート文書のタイトルが表示され、“Sub Topic:”に続いてそのスレッド中に現れるルート文書のタイトル以外のタイトルが表示される。そして、各タイトルは色分けされ、各タイトルと同じタイトル(参照記号を除く)を有する文書に対応するノードは、そのタイトルの色と同じ色で表示される。

【0096】これによって、ユーザは、スレッド全体の構造を把握しスレッド内の話題の推移を一目で把握することが可能となる。更に、各ノードはアンカーとして表示される。これにより、ユーザは、各ノードをマウス装置等によってクリックすることにより、そのノードに対応する文書に即座にアクセスすることができる。

【0097】図14は、表示装置306が実行するスレッドビューの表示動作を示す動作フローチャートである。まず、表示装置306は、図6のデータ構造を有するスレッドインデックス305のファイルから、1つのスレッドに対応する1つのエントリ(1行)のリストと、そのスレッドに含まれる文書数を、読み込む(ステップ1401)。例えば、図6のデータ構造を有するスレッドインデックス305において、ルート文書番号“001”に対応するリストとして、
(001 (002 003) (004 (005 006)))
が読み込まれ、文書数として“6”が読み込まれる。

【0098】次に、表示装置306は、読み込んだリストから、例えば図6の表の右側に示されるスレッドのツリー構造を復元する(ステップ1402)。このツリー構造を表現するために、表示装置306は、例えば図15に示されるような配列データを生成する。

【0099】次に、表示装置306は、読み込んだリストの各ノードを構成する文書番号毎に、その文書番号に対応する図5のデータ構造を有するメタインデックス304のエントリを抽出し、そのエントリから、作者名、行数、色番号、及びタイトルを抽出する（ステップ1403）。これらの抽出されたデータは、上記各ノードに対応付けて記憶される。

【0100】次に、表示装置306は、ステップ1401で読み込んだ文書数と、ステップ1403で抽出した各ノードの色番号とから、スレッドビューの先頭で表示される各タイトルの色を決定する（ステップ1404）。この動作は、各色番号に実際の色をマッピングする動作として実現される。

【0101】次に、表示装置306は、スレッドに含まれるルート文書のタイトルとその他のタイトルを、“Main Topic:”及び“Sub Topic:”に続けて表示するためのHTML文書を作成する。この場合に、各タイトルは、前述した構造解析部402が管理する図12に示されるカラーテーブルの上記スレッドに属する各エントリから順次読み出され、同時に順次読み出される各色番号からステップ1404で決定された各色が算出され、その各色での表示が順次指定される。各色は、HTML文書の色指定命令（<font color=タグ等）によって指定される。

【0102】最後に、表示装置306は、ステップ1402で復元したスレッドのツリー構造を示す配列データを構成する左端のノードの文書番号から順に処理することにより、そのツリー構造を表示するためのHTML文書を作成する（ステップ1406）。この場合、前述したように、表示装置306は、ステップ1403で抽出した各ノードの作者名、行数、及び色番号に基づいて、ツリー構造の各ノードの文書番号を、そのノードに対応する文書の作者名の先頭文字（2バイト）とその文書の行数とからなる表示データ、作者名【行数】に変換し、更に、その表示データをそのノードの色番号に対応する色で表示させるためのHTML文書データを生成する。色番号と実際の色との対応関係は、ステップ1404で決定された対応関係に従う。また、前述したように、表示装置306は、各ノードに対応する上記表示データの前後に、その接続関係に基づいて、“*”、“+”、“=”、“又は”、“.”等の記号を表示するためのHTML文書データを生成する。ここで、ツリー構造をそのままの形式で表示可能とするために、例えば、HTMLにおける制御用タグであるプリフォーマットタグ<PRE>が使用される。更に、上記ノード毎の表示データは、そのノードに対応する文書群データベース301内の文書データを表示するためのアプリケーションへのURLがHREF属性の値として指定されるアンカータグとして生成される。

【0103】続いて、表示装置306は、スレッドイン

デックス305のファイルから文書末尾（EOF）を検出するまで、上記ステップ1401～1406の一連の処理を繰り返し実行することにより、各スレッド毎のビューデータを作成する（ステップ1407→1401）。

【0104】表示装置306は、スレッドインデックス305のファイルから文書末尾を検出すると（ステップ1407の判定がYES）、最終的に得られたHTMLテーブル文書データをWebサーバに引き渡して、スレッドビューの表示動作を終了する。この結果、ユーザの端末のWebブラウザアプリケーション上に、図26に例示されるような形式で、スレッドビューが表示される。

【0105】次に、表示装置306が実現する発言者ビューの表示動作につき説明する。前述したように、発言者ビューにおいては、各文書のタイトルが、発言者（作者）毎に分類され、かつ発言者が発言の多い順にソートされ、同一発言者内では日付順で、図27に示される表示形態で、表示される。

【0106】図16は、表示装置306が実行する発言者ビューの表示動作を示す動作フローチャートである。表示装置306は、発言者ビューを実現するために、図17のデータ構造を有する作者配列データを使用する。そして、表示装置306は、発言者ビューの表示開始時に、この作者配列データを初期化する（ステップ1601）。

【0107】次に、表示装置306は、図5のデータ構造を有するメタインデックス304のファイルから1つのエントリのデータを読み込む（ステップ1602）。次に、表示装置306は、このエントリから抽出される作者名の作者が、作者配列データに含まれていない作者であるか否かを判定する（ステップ1603）。

【0108】表示装置306は、上記エントリから抽出される作者名の作者が、作者配列データに含まれていない作者である場合には、作者配列データに新しい作者項目を追加する（ステップ1603→1604）。表示装置306は、上記エントリから抽出される作者名の作者が、作者配列データに含まれている作者である場合には、ステップ1604の処理は実行しない。

【0109】次に、表示装置306は、作者配列データ中の該当する作者項目に、上記エントリから抽出される文書番号を登録する（ステップ1605）。続いて、表示装置306は、メタインデックス304のファイルから文書末尾（EOF）を検出するまで、上記ステップ1602～1605の一連の処理を繰り返し実行することにより、メタインデックス304に登録されている全ての文書番号を、作者別に作者配列データに登録する。

【0110】表示装置306は、メタインデックス304のファイルから文書末尾を検出すると（ステップ1606の判定がNO）、作者配列データ中の各作者項目

を、それぞれの項目に登録されている文書番号の数、即ち各作者毎の発言文書数に基づいてソートする（ステップ1607）。

【0111】続いて、表示装置306は、作者配列データ中の同一作者項目内で、文書番号を、それに対応するメタインデックス304中のエントリから抽出される日付に基づいてソートする（ステップ1608）。

【0112】最後に、表示装置306は、上記ステップ1607及び1608でのソートの結果得られる作者配列データの各作者項目毎に、作者名と、その項目内の各文書番号に対応するメタインデックス304中のエントリから抽出される日付及びタイトルを表示するためのHTMLテーブル文書データを生成し、それをWebサーバに引き渡して、発言者ビューの表示動作を終了する。この結果、ユーザの端末のWebブラウザアプリケーション上に、図27に例示されるようなテーブル形式で、発言者ビューが表示される。

【0113】ユーザは、発言者ビュー上で、発言者及び発言日付という観点から、文書集合（会議室）内の文書を参照することが可能となる。また、或る発言者の発言を時間を追って参照したり、会議室内で多くの発言をするリーダー的な発言者を一目で確認することができる。

表示装置306の他の表示態様

次に、上記各ビューの表示動作以外に表示装置306が実現する各表示動作の態様について説明する。

【0114】まず、表示装置306が実現する発言内容表示の動作につき説明する。前述したように、ユーザは、キーワードビューにおけるそれぞれのスレッド上のアンカー又はスレッドビューにおける各ノード上のアンカーを、マウス装置等でクリックすることにより、各スレッドのルート文書又は各ノードに対応する文書等に、即座にアクセスすることができる。

【0115】ユーザによってこれらの操作が実行された場合には、Webサーバから指示によって、表示装置306によって実行される図18に示される動作フローチャートの処理が例えばCGIとして起動される。この場合、この処理には、ユーザによって指定されたアンカータグに含まれる文書番号の情報が引き渡される。

【0116】この結果まず、表示装置306は、上記文書番号の情報を読み込んだ後（ステップ1801）、ヘッダ部に上記読み込んだ文書番号と同じ文書番号を含んでいる文書ファイルを読み込むまで、文書群データベース301からの文書ファイルの読み込みを行う（ステップ1802→1803→1802のループ）。

【0117】表示装置306は、ヘッダ部に上記読み込んだ文書番号と同じ文書番号を含んでいる文書ファイルを読み込むと（ステップ1803の判定がYES）、新しい文書のヘッダ部を読み込むまで、ステップ1804～1809のループにより、上記文書ファイルから1行ずつデータを読み込み、そのデータを1行分のHTML

文書データに変換し、そのHTML文書データをWebサーバに出力する（ステップ1808）。

【0118】この場合に、各行のデータが他の文書等へのURLを含んでいる場合には、表示装置306は、そのデータを上記URLがHREF属性の値として指定されるアンカータグに変換した上で出力する（ステップ1804→1805）。

【0119】この結果、ユーザは、発言内容の表示中のアンカーを更にマウス装置等によってクリックすることにより、更に他のリソースにジャンプすることができる。また、各行のデータが他の文書の行を引用したコメント行である場合には、表示装置306は、そのデータの色を変換するタグを追加した上で出力する（ステップ1806→1807）。

【0120】この結果、ユーザは、コメント行を一目で判別することができる。表示装置306は、該当する文書データの出力処理を終了すると、上記文書を含むスレッドのツリー構造を表示するHTML文書データを生成し出力して、発言内容表示の動作を終了する（ステップ1809→1810）。この処理は、前述した図14の動作フローチャートで示されるスレッドビューの表示動作と同様にして実現できる。

【0121】以上の表示動作の結果、ユーザの端末のWebブラウザアプリケーション上には、例えば図28に示されるように、表示画面の上半分に発言内容が表示され、表示画面の下半分にはその発言内容の文書を含むスレッドのツリー構造が表示される。なお、この表示画面には、図28に示されるように、キーワードビューやスレッドビューを表示させるためのアンカーや、検索を実行するためのアンカーと等を同時に表示させることもできる。

【0122】これらのビューの切替え機能により、例えば、キーワードビュー → スレッドビュー → 発言内容表示 → 発言者ビュー → 発言内容表示 → スレッドビュー → … というように、会議室内の文書（発言）をユーザの嗜好に応じて横断的に参照してゆることが可能となる。

【0123】次に、表示装置306が実現する作者別/日付別色分け表示の動作につき説明する。図19は、その動作を示す動作フローチャートである。まず、表示装置306は、メタインデックス304及びスレッドインデックス305に基づいて、図20(a)に示されるように作者項目毎に文書番号が分類された作者配列データと、図20(b)に示されるように日付項目毎に文書番号が分類された日付配列データとを予め作成する。これらの作成処理の詳細は省略するが、前述した図16の動作フローチャートと同様の処理によって実現できる。そして、作者配列データ中の各作者項目又は日付配列データ中の各日付項目に、それぞれ異なる色が割り当てられる。この色の割当ては、作者項目毎の作者の総数又は日

付項目毎の日付の総数から決定される。

【0124】次に、表示装置306は、ユーザの指定に基づく項目選択ボタン情報をWebサーバを経由して取得し、作者ボタンが押されたか日付選択ボタンが押されたかを判定する(ステップ1902、1904、図29参照)。

【0125】表示装置306は、作者ボタンが押されたと判定した場合には、図20(a)に示される作者配列データを参照することにより、スレッドツリーの表示データを作成し出力する(ステップ1902→1903)。この処理は、前述した図14の動作フローチャートと同様の処理によって実現されるが、この場合に、ツリーの各ノードは、そのノードに対応する作者名に対応する作者配列データ中の作者項目に割当てられている色で表示される。

【0126】一方、表示装置306は、日付選択ボタンが押されたと判定した場合には、図20(b)に示される日付配列データを参照することにより、スレッドツリーの表示データを作成し出力する(ステップ1904→1905)。この処理も、前述した図14の動作フローチャートと同様の処理によって実現されるが、この場合に、ツリーの各ノードは、そのノードに対応する日付に対応する日付配列データ中の日付項目に割当てられている色で表示される。

【0127】以上の表示動作の結果、例えばユーザが作者ボタンを押した場合には、ユーザの端末のWebブラウザアプリケーション上には、例えば図29に示されるように、作者別に色分けされたスレッドのツリーが表示され、ユーザは同一の作者の文書を一目で確認することができる。

【0128】次に、表示装置306が実現するスレッドビューを使った検索結果の強調表示の動作につき説明する。図21はその動作を示す動作フローチャートである。まず、表示装置306は、検索後入力フォーム画面を表示するためのHTML文書データを生成し出力する(ステップ2101)。この結果、ユーザの端末のWebブラウザアプリケーション上には、例えば図30に示されるような検索入力フォーム画面が表示される。ユーザは、この検索入力フォームに検索語を入力して検索の実行を指定する。

【0129】上記検索入力フォームに入力された検索語は、Webサーバを経由して文字列検索装置405(図4)に引き渡される。文字列検索装置405は、ユーザによる検索語の指定に基づいて、索引ファイル404を用いながら文書群データベース301内の指定されたスレッドを構成する各文書に対して全文検索を実行し、その検索語を含む文書番号を出力する(ステップ2102、2103)。

【0130】表示装置306は、上記検索語を含む文書番号を受け取ると、その文書番号を含むスレッドのツリ

ー構造を表示するHTML文書データを、前述した図14の動作フローチャートと同様の処理によって表示する。この場合に、表示装置306は、上記文書番号を含むノードの色を強調色に指定する(ステップ2104、2105)。

【0131】この結果、ユーザの端末のWebブラウザアプリケーション上において、例えば図31に示されるような検索結果に基づくスレッドビューの強調表示が実現される。これにより、ユーザは、スレッドの構造を把握しつつ、検索を実行することができる。

【0132】次に表示装置306が実現するキーワードビューを使った検索結果の強調表示の動作につき説明する。図22はその動作を示す動作フローチャートである。まず、表示装置306は、図21のステップ2101の場合と同様に、検索後入力フォーム画面を表示するためのHTML文書データを生成し出力する(ステップ2201)。ユーザは、この検索入力フォームに検索語を入力して検索の実行を指定する。

【0133】上記検索入力フォームに入力された検索語は、Webサーバを経由して文字列検索装置405(図4)に引き渡される。文字列検索装置405は、ユーザによる検索語の指定に基づいて、索引ファイル404を用いながら文書群データベース301内の指定された会議室を構成する各文書に対して全文検索を実行し、その検索語を含む文書番号を出力する(ステップ2202、2203)。

【0134】表示装置306は、上記検索語を含む文書番号を受け取ると、まず、図6のデータ構造を有するスレッドインデックス305を参照して、上記文書番号を含むエントリに対応するルート文書番号を抽出する(ステップ2204)。

【0135】続いて、表示装置306は、指定された会議室に関するキーワードビューを表示するHTML文書データを、前述した図13の動作フローチャートと同様の処理によって表示する。この場合に、表示装置306は、ステップ2204で抽出されたルート文書番号に対応するスレッドのタイトル又はその表示エリア全体の色を強調色に指定し、更に、表示されるキーワード中に検索語が含まれている場合には、そのキーワードも強調色に指定する(ステップ2205、2206、2207)。

【0136】この結果、ユーザの端末のWebブラウザアプリケーション上において、例えば図32に示されるような検索結果に基づくキーワードビューの強調表示が実現される。これにより、ユーザは、検索語を含むスレッドを一目で把握することができる。

【0137】なお、表示装置306は、検索結果の文書番号とそれに対応するタイトルを、例えば図33に示されるように羅列して表示するように構成することも可能である。

【0138】最後に、表示装置306が実現するサブトピック毎のキーワードビューの表示動作について説明する。前述したキーワードビューは、スレッド毎にキーワードを表示するものであった。これに対して、サブトピック毎のキーワードビューでは、1つのスレッド内のサブトピック毎に、キーワードを抽出して表示することができる。

【0139】この動作において、表示装置306は、図24のデータ構造を有するサブトピックインデックスを使用する。サブトピックインデックスは、図12に示されるカラーテーブルのデータ構造に対して、キーワードフィールドが追加されたデータ構造を有する。

【0140】サブトピックインデックスは、実質的には前述したカラーテーブルを置き換えるものであるため、サブトピックインデックスにおけるキーワードフィールド以外のフィールドの内容は、構造解析部402による前述した図8のステップ810の処理によって予め登録されている。この場合、図8のステップ810の処理の説明において前述したように、カラーテーブルであるサブトピックインデックスには、ルート文書番号毎（スレッド毎）に、それに含まれるルート文書のタイトルを示すメインピックと、それ以外の文書のタイトルを示すサブピックのそれぞれに対応するエントリが得られる。表示装置306は、この登録内容を利用する。

【0141】図23は、表示装置306が実現するサブトピックからのキーワード抽出の制御を示す動作フローチャートである。まず、表示装置306は、各スレッドについて、サブトピックインデックス内のそのスレッドに含まれる各エントリに登録されている文書番号に基づいて、メインピック及びサブピック単位で、それぞれに属する文書群を各結合文書ファイルにまとめ（ステップ2301）、その結果得られる各結合文書ファイルを内容解析部403（図4）に入力する（ステップ2302）。

【0142】内容解析部403は、各結合文書ファイル別にキーワードを抽出し、その結果を表示装置306に返す。表示装置306は、内容解析部403から返された各結合文書ファイル別のキーワードを、サブトピックインデックス内の上記各結合文書ファイルに対応するエントリのキーワードフィールドに登録する（ステップ2303）。

【0143】以上のようにして、各スレッドについて、メインピック及びサブピック単位で、それぞれに属する文書群からキーワードが抽出される。その後は、表示装置306は、サブトピックインデックスの内容に基づいて、ユーザにより指定されたスレッドに関して、そのスレッドのメインピック及びサブピック単位で、それぞれのタイトルとそれぞれに属するキーワードを表示するためのHTML文書データを生成し出力する。

【0144】この結果、ユーザの端末のWebブラウザ

アプリケーション上には、例えば、図34に示されるような形式で、サブトピック毎のキーワードビューが表示される。これにより、ユーザは、キーワードによるより精密なトピックの絞込みを行うことができる。

本発明の他の実施の形態（第2の実施の形態）

次に、本発明の他の実施の形態（以下、第2の実施の形態という）について説明する。

本発明の第2の実施の形態が実現する機能

まず、第2の実施の形態では、以下の3つの機能が実現される。

1. 狭い画面の中でのスレッドの全体構造の把握機能
スレッドのツリーが縮退（curtail）させられることにより、ある大きさの画面内でツリー構造の全体表示が可能となる。第2の実施の形態では、TTYキャラクタ端末上での表示を例に説明する。

【0145】キャラクタ端末では、1行に1ノードが表示されるため、n行の画面内にはn個のノードを描画することができる。描画されるノードの選択基準としては、以下のものがある。

【0146】・そのノードを参照している子ノードの個数。

・検索結果として得られる、そのノードを参照している子ノードの個数。

・ルートノード又は親ノードと異なるタイトルを持つノード。

2. スレッド内の話題の進行の推測機能

「質問—答え—お礼」といった特定の会話パターンが検出され、その情報が表示・検索に使用されることにより、効率の良い情報アクセスが可能となる。

【0147】より具体的には、第2の実施の形態では、文書の属性情報（タイトル、作者、参照関係）と、文書の内容を特徴づける特定の文章パターンが推測されることにより、スレッド内の話題パターンが抽出される。

3. 利用者の発言パターンの視覚化

ネットワークニュースでは、読んでいる人に対して発言する人の割合は非常に少ない。大きなスレッドであっても実は数人の人が論争しているだけという場合も少なくない。また、特定のニュースグループにおいて、有用な情報を発信する人が決まっている場合も多い。そこで、第2の実施の形態では、記事を投稿する利用者の観点から、ニュースやスレッドが整理されることにより、新たなビューが提供される。

【0148】より具体的には、第2の実施の形態では、ニュースグループ内の投稿履歴と、前述の話題推測機能に基づいて、利用者を観点とするビューが提供される。

本発明の第2の実施の形態の全体構成

図35は、本発明の第2の実施の形態の構成図である。

【0149】まず、検索フェーズの前の準備フェーズにおいては、以下の動作が実行される。処理装置3501内の文書取得部3502は、ネットワークを通じて、参

照関係のある文書群を取得し、二次記憶装置3503内に格納する。

【0150】内容推定部3504は、二次記憶装置3503に格納されている文書群の文書内容、文書付随情報、文書間の参照関係に基づいて、表示用インデックス3505を作成する。

【0151】検索エンジン3506は、二次記憶装置3503に格納されている文書群の文書内容に基づいて、検索用インデックス3507を作成する。例えばネットワークニュースサービスでは、文書が随時投稿されてゆく。そのため、上記の準備フェーズは、例えば一日に一度のように定期的に行われ、二次記憶装置3503には、常に最新の文書群が格納される。

【0152】検索フェーズの実行時には、以下の動作が実行される。利用者は、入力装置3509から入力指示を行う。入力される情報には、検索キーワードと、検索結果を表示させるためのビューの種類、ビューの表示領域の大きさが含まれる。

【0153】ビュー生成部3508は、入力装置3509からの入力指示に基づいて、検索エンジン3506を呼び出し、それに対して二次記憶装置3503に格納されている文書群の中から上記入力指示に対応する文書群を検索させる。

【0154】ビュー生成部3508は、表示用インデックス3505を利用して、検索エンジン3506が検索した文書群を表示するための結果ビューを作成し、それを表示装置3510に出力する。この場合に、後述するスレッド木の縮退処理が実行される。

【0155】以上の動作は、利用者との間の対話処理に基づいて実行される。つまり、利用者は、結果表示を見て、検索キーワードを追加又は変更し、或いは、結果ビューを切り替える。

表示用インデックス3505の構造

第2の実施の形態において、検索前の準備フェーズでは、以下の種類のインデックスが作成される。

1. ユーザインデックス

このインデックスは、ユーザの管理を行うためのインデックスであり、図36に示されるように、エントリ毎に、下記情報を保持する。

・ユーザID (User ID) : そのエントリに対応するユーザのID (キー) である。

【0156】・名前 : そのエントリに対応するユーザの名前である。

・略称 : そのエントリに対応するユーザの略称である。

・発言数 (回答数) : そのエントリに対応するユーザの、会議室内における発言の総数と、Q and Aパターンにおける回答文書の数である。

2. 文書インデックス

このインデックスは、文書毎の情報管理を行うためのインデックスであり、図37に示されるように、エントリ

毎に、下記情報を保持する。

・文書ID : そのエントリに対応する文書のID (キー) である。

【0157】・ユーザID (User ID) : そのエントリに対応する文書を作成したユーザのID (キー) である。

・タイトル : そのエントリに対応する文書のタイトルである。

【0158】・日付 : そのエントリに対応する文書の作成日である。

・参照子孫数 : そのエントリに対応する文書を参照する文書の総数である。

・ルートまでのパス : そのエントリに対応する文書が参照する先頭記事からその文書までのパスである。

【0159】・タイトルの識別番号 : そのエントリに対応する文書のタイトルが、その文書が含まれるスレッド (文書群) 中の何番目のタイトルであるかを示す番号である。

【0160】・記事種別 : そのエントリに対応する文書が、Q and Aパターンに含まれる場合に、その文書がQ (質問) 文書、A (答) 文書、又はT (お礼) 文書の何れにあたるかを示す情報である。

3. スレッドインデックス

このインデックスは、スレッド毎の情報管理を行うためのインデックスであり、図38に示されるように、エントリ毎に、下記情報を保持する。

・スレッドID : そのエントリに対応するスレッドのID (キー) である。

【0161】・スレッドの木構造 : そのエントリに対応するスレッド内の文書の参照関係を文書IDのリストで表現したものである。

・文書数 : そのエントリに対応するスレッド内の文書の総数である。

【0162】・作者数 : そのエントリに対応するスレッド内の文書の作者の数である。

・最多発言UID : そのエントリに対応するスレッド内で最も多く発言した作者のユーザIDである。

【0163】・内容リスト : そのエントリに対応するスレッドに含まれるQ and Aパターン、論争 (Discussion) パターン、又は雑談 (Chat) パターンのパターンIDのリストである。Q and AパターンのパターンIDであるQA_IDは、後述するQAインデックス内のいずれかのエントリに登録されている。DiscussionパターンのパターンIDであるDS_IDは、後述するDISCUSSインデックス内のいずれかのエントリに登録されている。ChatパターンのパターンIDであるCT_IDは、後述するCHATインデックス内のいずれかのエントリに登録されている。

4. QAインデックス

このインデックスは、Q and Aパターンの情報管理を行うためのインデックスであり、図39に示されるように、エントリ毎に、下記情報を保持する。

・QA_ID: そのエントリに対応するQ and AパターンのID (キー) である。

【0164】・Question: そのエントリに対応するQ and Aパターンを構成するQ (質問) 文書に対応する文書IDを格納するフィールドである。

・Answer: そのエントリに対応するQ and Aパターンを構成するA (答え) 文書群に対応する文書ID列を格納するフィールドである。

【0165】・Thanks: そのエントリに対応するQ and Aパターンを構成するT (お礼) 文書に対応する文書IDを格納するフィールドである。

・MaxAnswerUID: そのエントリに対応するQ and Aパターンを構成する各A (答え) 文書の作者、すなわち、そのエントリ内の「Answer」フィールドに登録されている文書ID列中の各文書IDに対応する文書の作者うち、もっとも登場回数が多い人 (又は人達) のユーザID (又はユーザID列) を格納するフィールドである。

【0166】・ThreadID: そのエントリに対応するQ and Aパターンが存在するスレッドのIDである。このスレッドIDは、スレッドインデックス内のいずれかのエントリに登録されている。

5. DISCUSSインデックス

このインデックスは、Discussionパターンの情報管理を行うためのインデックスであり、図40に示されるように、エントリ毎に、下記情報を保持する。

・DS_ID: そのエントリに対応するDiscussionパターンのID (キー) である。

【0167】・記事IDリスト: そのエントリに対応するDiscussionパターンを構成する文書群の文書ID列を格納するフィールドである。

・UID: そのエントリに対応するDiscussionパターンを構成する文書群のユーザID列である。

【0168】・ThreadID: そのエントリに対応するDiscussionパターンが存在するスレッドのIDである。このスレッドIDは、スレッドインデックス内のいずれかのエントリに登録されている。

6. CHATインデックス

このインデックスは、Chatパターンの情報管理を行うためのインデックスであり、図41に示されるように、エントリ毎に、下記情報を保持する。

・CT_ID: そのエントリに対応するChatパターンのID (キー) である。

【0169】・Chatリスト: そのエントリに対応するChatパターンを構成する文書群の文書ID列を格納するフィールドである。

・ThreadID: そのエントリに対応するChat

パターンが存在するスレッドのIDである。このスレッドIDは、スレッドインデックス内のいずれかのエントリに登録されている。

内容推定部3504の構成及び動作

図35に示される第2の実施の形態における内容推定部3504の動作について、以下に詳細に説明する。

【0170】前述したように、内容推定部3504は、二次記憶装置3503に格納されている文書群の文書内容、文書付随情報、文書間の参照関係に基づいて、表示用インデックス3505であるユーザインデックス、文書インデックス、スレッドインデックス、QAインデックス、DISCUSSインデックス、及びCHATインデックスを作成する。

【0171】図42は、内容推定部3504が実行する動作を示す動作フローチャートである。まず、図37に示されるデータ構成を有する文書インデックスと図38に示されるデータ構成を有するスレッドインデックスが作成される (ステップ4201)。これらの詳細は省略するが、基本的に、前述した図8及び図9に示される動作フローチャートと同様の動作によって実現できる。この場合には、前述したメタインデックスが文書インデックスに対応する。このとき同時に、各文書中に現れる作成ユーザ名とユーザID、略称、及び発言数 (回答数) を対応づけるための図36に示されるデータ構成を有するユーザインデックスも作成される。

【0172】次に、スレッドインデックス内の各エントリが参照されることにより、各エントリに対応するスレッド文書群が読み込まれ (ステップ4202)、全てのエントリに対するスレッド文書群の処理が終了したと判定されるまで (ステップ4206)、読み込まれたスレッド文書群毎に、Q and Aパターンの判定処理 (ステップ4203)、Discussionパターンの判定処理 (ステップ4204)、及びChatパターンの判定処理 (ステップ4205) が実行される。

【0173】図43は、図42のステップ4203のQ and Aパターンの判定処理の動作フローチャートである。この動作フローチャートでは、スレッド文書群内の各参照パス毎に、Q and Aパターンが推測される。

【0174】まず、スレッドインデックス内の該当エントリの「スレッドの木構造」フィールドが参照されることによって、リーフ文書 (パスの末端の文書) に対応する文書IDが1つ選択される (ステップ4301)。

【0175】次に、文書インデックスにおいて、ステップ4301で選択された文書IDを「文書ID」フィールドに含むエントリ内の「ルートまでのパス」フィールドから、下記条件を満たす文書IDが検索される (ステップ4302)。

(条件) 文書インデックスにおいて、その文書IDを「文書ID」フィールドに含むエントリ内の「記事種別」フィールドが未登録である。続いて、上記条件を満た

たす文書IDが見つかったか否かが判定される（ステップ4303）。

【0176】上記条件を満たす文書ID（以下、処理文書IDという）が見つかりステップ4303の判定がYESとなった場合には、その処理文書IDに対応する文書が二次記憶装置3503（図35）から読み出され、その文書中に、図44に示されるような、センテンスパターンが存在するか否かが判定される（ステップ4304）。

【0177】ステップ4304の判定がNOならば、ステップ4308にジャンプする。ステップ4304の判定がYESならば、文書インデックスのステップ4302で参照されたエントリ内の「ルートまでのパス」フィールドに登録されている文書IDのうち、下記条件を満たす文書IDが存在するか否かが判定される（ステップ4305）。

（条件）その文書IDに対応する文書は、処理文書IDの作者によって作成されたものであって、かつその文書IDは、図39に示されるデータ構成を有するQAインデックス内のいずれかのエントリ内の「Thanks」フィールドに登録されている。ステップ4305の判定がYESなら、ステップ4305で参照されたQAインデックス内のエントリの「Question」フィールドに、処理文書IDが追加される。また、図37に示されるデータ構成を有する文書インデックスにおいて、処理文書IDを「文書ID」フィールドに含むエントリ内の「記事種別」フィールドに、記号「Q」が追加される（ステップ4306）。

【0178】更に、文書インデックスのステップ4302で参照されたエントリ内の「ルートまでのパス」フィールドに登録されている文書ID群のうち、ステップ4305で参照されたQAインデックス内のエントリの「Question」フィールドに登録された処理文書IDとそのエントリの「Thanks」フィールドに登録された文書IDに挟まれた文書ID群が、そのエントリの「Answer」フィールドに追加される。また、図37に示されるデータ構成を有する文書インデックスにおいて、上記登録が行われた各文書IDを各「文書ID」フィールドに含む各エントリ内の「記事種別」フィールドに、それぞれ記号「A」が追加される（ステップ4306）。

【0179】一方、ステップ4306の判定がNOなら、QAインデックスにおいて、{(そのインデックス内のQA_IDの最大値)+1}の値を「QA_ID」フィールドの値として有するエントリが作成され、そのエントリ内の「Question」フィールドに、処理文書IDが登録される。また、図37に示されるデータ構成を有する文書インデックスにおいて、処理文書IDを「文書ID」フィールドに含むエントリ内の「記事種別」フィールドに、記号「Q」が登録される（ステップ

4307）。

【0180】上記ステップ4306又は4307の処理の後、又はステップ4304の判定がNOとなった場合には、二次記憶装置3503から読み出されている処理文書IDに対応する文書中に、図45に示されるような、センテンスパターンが存在するか否かが判定される（ステップ4308）。

【0181】ステップ4308の判定がNOならば、ステップ4302に戻る。ステップ4308の判定がYESならば、文書インデックスのステップ4302で参照されたエントリ内の「ルートまでのパス」フィールドに登録されている文書IDのうち、下記条件を満たす文書IDが存在するか否かが判定される（ステップ4309）。

（条件）その文書IDに対応する文書は、処理文書IDの作者によって作成されたものであって、かつその文書IDは、図39に示されるデータ構成を有するQAインデックス内のいずれかのエントリ内の「Question」フィールドに登録されている。ステップ4309の判定がYESなら、ステップ4309で参照されたQAインデックス内のエントリの「Thanks」フィールドに、処理文書IDが追加される。また、図37に示されるデータ構成を有する文書インデックスにおいて、処理文書IDを「文書ID」フィールドに含むエントリ内の「記事種別」フィールドに、記号「T」が追加される（ステップ4310）。

【0182】更に、文書インデックスのステップ4302で参照されたエントリ内の「ルートまでのパス」フィールドに登録されている文書ID群のうち、ステップ4309で参照されたQAインデックス内のエントリの「Thanks」フィールドに登録された処理文書IDとそのエントリの「Question」フィールドに登録された文書IDに挟まれた文書ID群が、そのエントリの「Answer」フィールドに追加される。また、図37に示されるデータ構成を有する文書インデックスにおいて、上記登録が行われた各文書IDを各「文書ID」フィールドに含む各エントリ内の「記事種別」フィールドに、それぞれ記号「A」が追加される（ステップ4310）。

【0183】一方、ステップ4309の判定がNOなら、QAインデックスにおいて、{(そのインデックス内のQA_IDの最大値)+1}の値を「QA_ID」フィールドの値として有するエントリが作成され、そのエントリ内の「Thanks」フィールドに、処理文書IDが登録される。また、図37に示されるデータ構成を有する文書インデックスにおいて、処理文書IDを「文書ID」フィールドに含むエントリ内の「記事種別」フィールドに、記号「T」が登録される（ステップ4311）。

【0184】上記ステップ4310又は4311の処理

の後、ステップ4302に戻り、次の文書IDの検索が実行される。上記ステップ4302～4311の処理が繰り返された結果、ステップ4303で、ステップ4302における条件を満たす文書IDが見つからなかったと判定された場合には、スレッドインデックス内の現在処理中のエントリの「スレッドの木構造」フィールドが参照されることによって、全てのリーフ文書に対応する文書IDに対する処理が試行されたか否かが判定される（ステップ4312）。

【0185】全てのリーフ文書に対応する文書IDに対する処理が試行されてはならずステップ4312の判定がNOの場合には、ステップ4301に戻り、次のパスに対応する話題パターンの推測処理が繰り返される。

【0186】全てのリーフ文書に対応する文書IDに対する処理が試行されステップ4312の判定がYESとなった場合には、図42のステップ4203のQ and Aパターンの判定処理を終了する。

【0187】図46及び図47に、上述のQ and Aパターンの判定処理によって抽出されるスレッド構造とそれに対応する文書群の例を示す。なお、文書インデックスの「記事種別」フィールドに記号「A」が付与されたエントリの文書IDに対応する文書の作者について、それに対応するユーザインデックス（図36参照）のエントリが参照され、そのエントリ内の「発言数（回答数）」フィールドの内容が更新される。

【0188】図48は、図42のステップ4204のDiscussionパターンの判定処理の動作フローチャートである。この動作フローチャートでは、スレッド文書群内の各参照パス毎に、Discussionパターンが推測される。

【0189】まず、スレッドインデックス内の該当エントリの「スレッドの木構造」フィールドが参照されることによって、リーフ文書（パスの末端の文書）に対応する文書IDが検索される（ステップ4801）。

【0190】次に、上記検索の結果、全てのリーフ文書に対応する文書IDに対する処理が試行されたか否かが判定される（ステップ4802）。全てのリーフ文書に対応する文書IDに対する処理が試行されてはならずステップ4802の判定がNOの場合には、文書インデックスにおいて、ステップ4801で検索された文書IDを「文書ID」フィールドに含むエントリ内の「ルートまでのパス」フィールドが参照され、上記リーフ文書に対応する文書IDからルート文書までの長さ（文書IDの数）が6以上であるか否かが判定される（ステップ4803）。

【0191】上記長さが6以上ではなくステップ4803の判定がNOの場合には、その参照パスの話題パターンはDiscussionパターンではないと推測され、ステップ4801に戻って次のリーフ文書に対する処理が実行される。

【0192】上記長さが6以上であってステップ4803の判定がYESの場合には、ステップ4803で参照された「ルートまでのパス」フィールドに含まれる文書ID群に対応する文書群において、相異なるユーザIDの数がカウントされる（ステップ4804）。

【0193】次に、{上記「ルートまでのパス」フィールドに含まれる文書IDの数（総文書数）}に対する{上記相異なるユーザIDの数}の割合が、0.3より小さいか否かが判定される（ステップ4805）。

【0194】この判定がNOの場合には、特定の少数のユーザによる論争が行われてはいないと推測され、ステップ4801に戻って次のリーフ文書に対する処理が実行される。

【0195】一方、ステップ4805の判定がYESの場合には、特定の少数のユーザによる論争が行われていると推測され、図40に示されるデータ構成を有するDISCUSSインデックスにおいて、{(そのインデックス内のDS_IDの最大値)+1}の値を「DS_ID」フィールドの値として有するエントリが作成される。そして、そのエントリ内の「記事ID」フィールドに、ステップ4803で参照された「ルートまでのパス」フィールドに含まれる文書ID群がリストとして登録され、その登録内容に基づいて、「UID」フィールド及び「ThreadID」フィールドの内容が登録される。また、図37に示されるデータ構成を有する文書インデックスにおいて、上記各文書ID群を各「文書ID」フィールドに含む各エントリ内の「記事種別」フィールドに、記号「D」が登録される（ステップ4806）。その後、ステップ4801に戻って次のリーフ文書に対する処理が実行される。

【0196】全てのリーフ文書に対応する文書IDに対する処理が試行されステップ4802の判定がYESとなった場合には、図42のステップ4204のDiscussionパターンの判定処理を終了する。

【0197】図49に、上述のDiscussionパターンの判定処理によって抽出されるスレッド構造の例を示す。少数のユーザの頭文字のみが多く現れていることがわかり、このスレッドにおいては論争が行われていると推測できる。

【0198】図50は、図42のステップ4205のChatパターンの判定処理の動作フローチャートである。この動作フローチャートでは、スレッド文書群内の各参照パス毎に、Chatパターンが推測される。

【0199】まず、スレッドインデックス内の該当エントリの「スレッドの木構造」フィールドが参照されることによって、リーフ文書（パスの末端の文書）に対応する文書IDが検索される（ステップ5001）。

【0200】次に、上記検索の結果、全てのリーフ文書に対応する文書IDに対する処理が試行されたか否かが判定される（ステップ5002）。全てのリーフ文書に

対応する文書IDに対する処理が試行されてはおらずステップ5002の判定がNOの場合には、文書インデックスにおいて、ステップ5001で検索された文書IDを「文書ID」フィールドに含むエントリ内の「ルートまでのパス」フィールドが参照され、上記リーフ文書に対応する文書IDからルート文書までの長さ（文書IDの数）が6以上であるか否かが判定される（ステップ5003）。

【0201】上記長さが6以上ではなくステップ5003の判定がNOの場合には、その参照パスの話題パターンはChatパターンではないと推測され、ステップ5001に戻って次のリーフ文書に対する処理が実行される。

【0202】上記長さが6以上であってステップ5003の判定がYESの場合には、ステップ5003で参照された「ルートまでのパス」フィールドに含まれる文書ID群に対応する文書群において、相異なるユーザIDの数がカウントされる（ステップ5004）。

【0203】次に、{上記「ルートまでのパス」フィールドに含まれる文書IDの数（総文書数）}に対する{上記相異なるユーザIDの数}の割合が、0.6より大きいかが判定される（ステップ5005）。

【0204】この判定がNOの場合には、多数のユーザによる雑談（チャット）が行われてはいないと推測され、ステップ5001に戻って次のリーフ文書に対する処理が実行される。

【0205】一方、ステップ5005の判定がYESの場合には、多数のユーザによる雑談が行われていると推測され、図41に示されるデータ構成を有するCHATインデックスにおいて、{(そのインデックス内のCT_IDの最大値)+1}の値を「CT_ID」フィールドの値として有するエントリが作成される。そして、そのエントリ内の「Chatリスト」フィールドに、ステップ5003で参照された「ルートまでのパス」フィールドに含まれる文書ID群がリストとして登録され、その登録内容に基づいて、「UID」フィールド及び「ThreadID」フィールドの内容が登録される（ステップ4806）。その後、ステップ5001に戻って次のリーフ文書に対する処理が実行される。

【0206】全てのリーフ文書に対応する文書IDに対する処理が試行されステップ5002の判定がYESとなった場合には、図42のステップ4205のChatパターンの判定処理を終了する。

【0207】図51に、上述のChatパターンの判定処理によって抽出されるスレッド構造の例を示す。多数のユーザの頭文字が雑多に現れていることがわかり、このスレッドにおいては雑談が行われていると推測できる。

スレッド木の縮退処理の原理

次に、第2の実施の形態におけるスレッド木の縮退処理

の原理について説明する。

【0208】図52及び図53は、スレッド構造の表示例を示す図である。図52は、従来の伝統的なニュースリーダにおける表示例、図53は、前述した図14のスレッドビューの表示処理に基づく表示例である。

【0209】わずか17本の記事からなるスレッドにおいても、図52に示されるように行数が増えたり、図53に示されるように横方向にはみだしたりして、スレッド全体を見るのに画面のスクロールが必要となり、全体構造の把握が難しいことがわかる。

【0210】図54は、スレッド内の上位n個の子孫ノード（図54の例ではn=6）に対する表示例である。行頭の“+”記号は子ノードが省略されていることを表し、行末のかっこ付き数字は子孫ノードの個数（図37に示される文書インデックスにおける参照子孫数）を表す。画面やウィンドウサイズに合わせて値nを調整することにより、必要な部分のみを表示することが可能となる。

【0211】図55は、子孫ノードのうち、同一タイトル（先頭の“Re:”は除く）を持つノードが省略された表示例である。行頭の“+”記号は子ノードが省略されていることを表す。同一スレッドの文書においては、デフォルトではタイトルは、親ノードと同じであるか、又は、最初にフォローを表す“Re:”が付加されるかである。作者が意図的にタイトルを変えたというのは、そこで話題が変わったことを明示している。このビューにより、スレッド内にどのような話題の変化があったかを容易に把握することができる。

【0212】図56は、10/1から10/5という時間区間に作成された文書のスレッド構造の表示例である。ノードは、作者のイニシャルを表す。このビューでは、スレッドの時間的展開と一定区間内の情報だけを見ることができる。また、パソコンの画面上でスケジューラなどの時間的情報のあるアプリケーションと並べて見ることによって、自分のスケジュールや世の中の出来事と関連づけて文書情報を見ることができる。

【0213】図57は、スレッドの作者をノードとしたグラフ構造である。ノードは、作者のイニシャルである。二重丸で表されたノードは、スレッドの最初の記事の発言者を表す。リンクの濃さにより作者間のやりとりの回数表される。更に、図38に示されるスレッドインデックス中の該当エントリの最多発言UIDに登録されているユーザIDに対応するユーザのノードは、例えば強調表示される。このスレッドが、小川さん（小）とパーティバルさん（パ）とのやりとりが中心であることが容易に理解できる。

検索フェーズの実行時の動作

次に、上述のスレッド木の縮退処理を含む検索フェーズの実行時の動作について説明する。

【0214】検索時には、利用者は、入力装置3509

から検索要求を指示する。図58は、検索要求の入力画面である。入力項目としては、下記に示されるものがある。

・探したい記事に含まれるキーワード列（必須）。

【0215】・探したい記事が含んではいけないキーワード列。

・検索対象の記事の種別として、全ての記事か、Q and Aパターンに相当する記事だけか。省略時は全ての記事。

【0216】・検索対象の記事の日付として、全区間か、一ヶ月以内か、一週間以内か。省略時は、全区間。図58に示される入力画面の下部には、検索前の準備フェーズにおいて記事が二次記憶装置3503に格納（ダウンロード）された最新の日時が表示されている。

【0217】検索結果としては、下記に示されるものがある。

・スレッド一覧（図60参照）。

【0218】・スレッド構造表示（参照数による縮退表示、同一タイトルによる縮退表示を含む）（図61参照）。

・時間区間スレッド表示（図62参照）。

【0219】・Q and A対照表示。

・作者ノードグラフ表示（図57参照）。

・作者投稿一覧表示。

【0220】・記事本文表示。

これらの表示画面は、図59に示されるように相互に切り替えることができる。これらの表示画面のうち、代表的なものについて以下に説明する。

出力結果1：スレッド一覧

例えば、検索キーワードとして「エンジン」が入力された場合、図60に示されるようなスレッド一覧画面が表示される。図60で表示される検索結果は、下記に示されるものである。

・スレッドのトップ記事のタイトル。

【0221】・作者の名前。

・日付。

・サイズ（スレッドの記事数、全体の記事サイズ）。

【0222】・スレッドの内容（QA：Q and Aパターン、DC：Discussionパターン、CT：Chatパターン）。

検索時にはスレッドのサイズに基づくソーティング処理が実行され、上位10スレッドが表示される。ユーザが「次の10スレッド」をクリックすると、次の10スレッドが表示される。

【0223】また、結果が多い場合には、更にキーワードを追加することにより絞り込み検索を実行させることも可能である。他の画面へは、次の方法で移動することができる。

・タイトルをクリックすると、スレッド構造が表示される。

【0224】・作者の名前をクリックすると、作者ノードグラフが表示される。

・日付をクリックすると、時間区間スレッド表示が表示される。

・スレッドの内容のQAをクリックすると、QAの対が表示される。

出力結果2：スレッド構造表示

図61は、スレッド構造の表示例である。

【0225】図54に示されるように、スレッド構造が、参照ノード数に基づいて縮退された木構造として表示される。表示領域の行数（縦方向の長さ）に応じて、参照ノード数の少ないノードは省略して表示される。ノードの表示内容は、下記のとおりである。表示領域の桁数（横方向の長さ）に応じて、各ノードにおいて表示される項目も適宜省略される。

・行頭の“+”記号は、省略された子ノードがある場合に付加される。

【0226】・ユーザが入力したキーワードを含む記事は、タイトルと作者部分が強調表示される（図61では、矩形によって囲まれた部分）。

・記事タイトル。フォロー記事には“Re:”記号が付加される。

【0227】・記事の作者名。

・記事の種別。内容推定部3504（図35）によって推定された話題パターンに応じて、Q（質問）、A（答）、D（論争）が付加される。

【0228】・自ノードの子孫ノードの数。“+”記号が付加されたノードに対してのみ付加される。

また、このスレッド内において、更にキーワードを指定して絞り込み検索を実行することも可能である。

【0229】他の画面へは、次の方法で移動することができる。

・タイトルをクリックすると、記事本文が表示される。

【0230】・作者の名前をクリックすると、作者ノードグラフが表示される。

・「タイトル一覧」をクリックすると、スレッド構造が、同一タイトルに基づいて縮退された木構造として表示される（図55参照）。

出力結果3：時間区間スレッド表示

図62(a)は、一定の時間区間におけるスレッドの表示例である。キーワードが含まれる記事は黒丸によって、そうでない記事は灰色の丸によって表示されている。また、作者のイニシャルが各ノードの下に付加される。

【0231】この画面では、日付の表示区間、日付の縦横表示、ウィンドウのサイズ、セルの幅などが可変である。そこで、例えば他のスケジューラとサイズを合わせることが可能である。例えば、図62(b)は、他のスケジューラであり、それと図62(a)に示される時間区間スレッド表示とで、各セル幅が合わせられている。

【0232】他の画面へは、次の方法で移動することが

できる。

・黒丸又は灰色の丸をクリックすると、記事本文が表示される。

【0233】・作者のイニシャルをクリックすると、作者ノードグラフが表示される。

出力結果4：QA対表示

QA対表示とは、内容推定部3504（図35）によって推測されたQ and Aパターンに対応する質問と回答の対が、テーブルとして表示されたものである。テーブルの一行には、下記の情報が表示される。

・タイトル。

【0234】・質問者。

・回答者（複数）。

他の画面へは、次の方法で移動することができる。

・タイトルをクリックすると、スレッド構造が表示される。

【0235】・作者の名前をクリックすると、作者ノードグラフが表示される。

出力結果5：作者ノードグラフ

作者ノードグラフは、そのスレッド内の各記事の作者間の会話関係がグラフ化されたものである。前述の図57がその表示例である。

【0236】他の画面へは、次の方法で移動することができる。

・作者をクリックすると、その作者の投稿一覧が表示される。

【0237】・リンクをクリックすると、スレッド構造が表示される。

出力結果6：作者の投稿一覧

作者の投稿一覧は、各作者が投稿した記事の一覧を見るための画面である。日付、タイトル、記事の種別（Q、A、D）が日付順に表示される。

【0238】他の画面へは、次の方法で移動することができる。

・タイトルをクリックすると、記事本文が表示される。

出力結果7：記事本文

これは、記事の本文である。他に、作者名、タイトル、日付、親記事へのリンクが表示される。

【0239】他の画面へは、次の方法で移動することができる。

・タイトルをクリックすると、スレッド構造が表示される。

【0240】・日付をクリックすると、時間区間スレッド表示が表示される。

・作者の名前をクリックすると、作者ノードグラフが表示される。

時間区間スレッド表示の動作

図6.3は、ビュー生成部3508（図35）が実行する時間区間スレッド表示の動作フローチャートである。

【0241】まず、図38に示されるデータ構成を有す

るスレッドインデックスにおいて、表示対象スレッドに対応するエントリ内の「スレッドの木構造」フィールドが参照されることにより、そのスレッドに含まれる文書IDが1つ選択される（ステップ6301）。

【0242】次に、図37に示されるデータ構成を有する文書インデックスにおいて、上記選択された文書IDに対応するエントリ内の「日付」フィールドが参照され、その日付が図64に示されるデータ構成を有するカレンダーインデックスに登録される（ステップ6302）。

【0243】次に、スレッドインデックスの表示対象スレッドに対応するエントリ内の「スレッドの木構造」フィールド内の全ての文書IDに対する処理が試行されたか否かが判定される（ステップ6303）。

【0244】全ての文書IDに対する処理が試行されてはおらずステップ6303の判定がNOの場合には、ステップ6301に戻り、次の文書IDに対する処理が繰り返される。

【0245】全ての文書IDに対する処理が試行されステップ6303の判定がYESとなった場合には、図64に示されるデータ構成を有するカレンダーインデックスが参照されることにより、カレンダーに文書ノードがマッピングされる。参照関係のエッジは、スレッドインデックスを参照して表示される。

作者ノードグラフの表示動作

図65は、ビュー生成部3508（図35）が実行する作者ノードグラフの表示動作を示す動作フローチャートである。

【0246】まず、図38に示されるデータ構成を有するスレッドインデックスにおいて、表示対象スレッドに対応するエントリ内の「スレッドの木構造」フィールドが参照されることにより、そのスレッドに含まれる文書IDが1つ選択される。次に、図37に示されるデータ構成を有する文書インデックスと図36に示されるデータ構成を有するユーザインデックスとが参照されることにより、上記選択された文書IDに対応する文書の親文書（親発言）のユーザIDが取得される（ステップ6501）。

【0247】次に、図66に示されるデータ構成を有する発言者配列内に、上記親子関係に対応するエントリが存在するか否かが判定される（ステップ6502）。そのエントリが存在するなら、ステップ6504の処理に進む。

【0248】そのエントリが存在しないなら、発言者配列の横軸又は縦軸のエントリが追加される（ステップ6503）。その後、上記エントリの数字が1だけインクリメントされる（ステップ6504）。

【0249】次に、スレッドインデックスの表示対象スレッドに対応するエントリ内の「スレッドの木構造」フィールド内の全ての文書IDに対する処理が試行された

か否かが判定される（ステップ6505）。

【0250】全ての文書IDに対する処理が試行されてはおらずステップ6505の判定がNOの場合には、ステップ6501に戻り、次の文書IDに対する処理が繰り返される。

【0251】全ての文書IDに対する処理が試行されステップ6505の判定がYESとなった場合には、図67に示されるように、図66に示されるデータ構成を有する発言者配列のエントリの数だけノードが描画され、親から子供に向かって“→”線が描画される。この線の太さは、その親子間の会話の度数に応じて決定される（ステップ6506）。

【0252】以上説明した第2の実施の形態において、スレッドの木構造が縮退されることにより、画面の表示範囲に応じたスレッドの表示可能となる。また、自動的に推定された話題と共に検索結果が表示されるため、検索結果のスレッド数が多い場合でも、利用者は検索結果の概要を容易に把握することが可能となる。

【0253】更に、スレッド中の文書量が多くても、同じ作者が何度も投稿している場合がある。作者を中心に見せるビューが提供されることにより、スレッド内のキーパーソンが把握可能となるだけでなく、スレッドの全体構造もコンパクトに表示することが可能となる。

各実施の形態を実現するプログラムが記録された記録媒体についての補足

本発明は、計算機により使用されたときに、上述の本発明の各実施の形態の各構成によって実現される機能と同様の機能を計算機に行わせるための計算機読出し可能記憶媒体として構成することもできる。

【0254】この場合に、図68に示されるように、例えばフロッピーディスク、CD-ROMディスク、光ディスク、リムーバブルハードディスク等の可搬型記憶媒体6802や、ネットワーク回線6803経由で、本発明の好適実施例の各種機能を実現するプログラムが、コンピュータ6801の本体6804内のメモリ（RAM又はハードディスク等）6805にロードされて、実行される。

【0255】

【発明の効果】本発明の第1の態様の構成によれば、キーワードを頼りにして、雑多な文書集合の中から必要な文書が含まれているであろう文書群を正確かつ容易に見つけ出すことが可能となる。

【0256】本発明の第2の態様の構成によれば、検索指定に該当する文書群を一目で把握することが可能となる。本発明の第3の態様の構成によれば、文書群全体の構造を容易に把握することが可能となる。

【0257】本発明の第4の態様の構成によれば、文書群全体の話題の推移を容易に把握することが可能となる。本発明の第5の態様の構成によれば、スレッドの構造を把握しつつ、検索を実行することが可能となる。

【0258】本発明の第6の態様の構成によれば、キーワードによるより精密なトピックの絞込みを行うことが可能となる。また、本発明によれば、スレッドの木構造が縮退されることにより、画面の表示範囲に応じたスレッドの表示可能となる。

【0259】更に、本発明によれば、自動的に推定された話題と共に検索結果が表示されるため、検索結果のスレッド数が多い場合でも、利用者は検索結果の概要を容易に把握することが可能となる。

【0260】加えて、本発明によれば、スレッド中の文書量が多くても、同じ作者が何度も投稿している場合がある。作者を中心に見せるビューが提供されることにより、スレッド内のキーパーソンが把握可能となるだけでなく、スレッドの全体構造もコンパクトに表示することが可能となる。

【0261】このように、本発明によれば、文書群に対して種々の観点からアクセスすることが可能となる。

【図面の簡単な説明】

【図1】参照関係を有する文書群の例を示す図である。

【図2】文書のデータ構造の例を示す図である。

【図3】本発明の実施の形態のシステム構成図（その1）である。

【図4】本発明の実施の形態のシステム構成図（その2）である。

【図5】メタインデックスのデータ構造を示す図である。

【図6】スレッドインデックスのデータ構造を示す図である。

【図7】索引ファイルのデータ構造を示す図である。

【図8】書式解析部と構造解析部の動作フローチャートである。

【図9】文書番号のスレッドインデックスへの登録の動作フローチャートである。

【図10】文書番号のスレッドインデックスへの登録の動作説明図である。

【図11】色番号登録の動作フローチャートである。

【図12】カラーテーブルの例を示す図である。

【図13】キーワードビューの動作フローチャートである。

【図14】スレッドビューの動作フローチャートである。

【図15】スレッドビューの制御用配列の例を示す図である。

【図16】発言者ビューの動作フローチャートである。

【図17】発言者ビューの制御用配列の例を示す図である。

【図18】発言内容表示の動作フローチャートである。

【図19】作者別／日付別色分け表示の動作フローチャートである。

【図20】作者別／日付別色分け表示用配列の例を示す

図である。

【図21】スレッドビューを使った検索結果の強調表示の動作フローチャートである。

【図22】キーワードビューを使った検索結果の強調表示の動作フローチャートである。

【図23】サブトピックからのキーワード抽出の制御動作フローチャートである。

【図24】サブトピックインデックスの例を示す図である。

【図25】キーワードビューの表示例を示す図である。

【図26】スレッドビューの表示例を示す図である。

【図27】発言者ビューの表示例を示す図である。

【図28】発言内容の表示例を示す図である。

【図29】スレッドビューを用いた文書属性「作者」の強調（色別）表示の例を示す図である。

【図30】会議室内の全発言の検索（入力）表示例を示す図である。

【図31】スレッドビューを用いた文字列「プロトコル」を含むノードの強調表示の例を示す図である。

【図32】キーワードビューを用いた文字列「プロトコル」を含むスレッドの強調表示の例を示す図である。

【図33】会議室内の全発言の検索（結果出力）表示例を示す図である。

【図34】サブトピックから抽出したキーワードの表示例を示す図である。

【図35】本発明の他の実施の形態（第2の実施の形態）の構成図である。

【図36】ユーザインデックスの構成図である。

【図37】文書インデックスの構成図である。

【図38】スレッドインデックスの構成図である。

【図39】QAインデックスの構成図である。

【図40】DISCUSSインデックスの構成図である。

【図41】CHATインデックスの構成図である。

【図42】内容推定部の動作フローチャートである。

【図43】Q and Aパターン判定処理の動作フローチャートである。

【図44】Q（質問）文書に含まれるパターンの例を示す図である。

【図45】お礼文書に含まれるパターンの例を示す図である。

【図46】Q and Aパターンの判定処理が推測するスレッド構造の例を示す図である。

【図47】Q and Aパターンの判定処理が推測するスレッド文書群の例を示す図である。

【図48】論争パターンの判定処理の動作フローチャートである。

【図49】Discussionパターンの判定処理が推測するスレッド構造の例を示す図である。

【図50】雑談パターンの判定処理の動作フローチャー

トである。

【図51】Chatパターンの判定処理が推測するスレッド構造の例を示す図である。

【図52】オリジナルのスレッド構造の表示例を示す図である。

【図53】オリジナルのスレッド構造の表示例を示す図である。

【図54】参照ノード数により縮退したスレッド構造の表示例を示す図である。

【図55】同一タイトル文書を縮退したスレッド構造の表示例を示す図である。

【図56】時間区間で取り出したスレッド構造の表示例を示す図である。

【図57】作者をノードとしたグラフ構造を示す図である。

【図58】ネットワークニュース検索システムの表示例を示す図である。

【図59】検索画面の一覧を示す図である。

【図60】出力結果1：スレッド一覧を示す図である。

【図61】出力結果2：スレッド構造表示を示す図である。

【図62】出力結果3：時間区間スレッド表示を示す図である。

【図63】時間区間スレッド表示の作成フローを示す図である。

【図64】カレンダーインデックスのデータ構成図である。

【図65】作者ノードグラフの作成フローを示す図である。

【図66】発言者配列のデータ構成図である。

【図67】作者ノードグラフの作成説明図である。

【図68】本実施の形態を実現するプログラムが記録された記録媒体の説明図である。

【符号の説明】

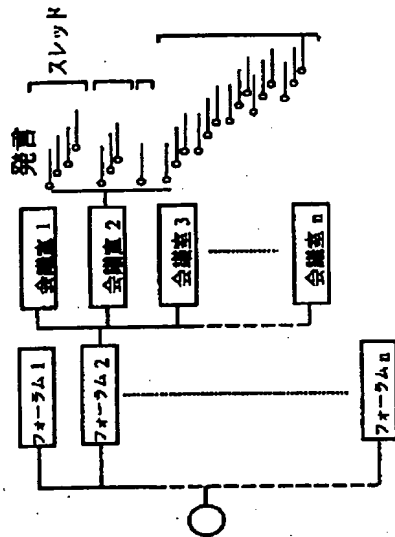
301	文書群データベース
302	文書群解析装置
303	集計装置
304	メタインデックス
305	スレッドインデックス
306	表示装置
401	書式解析部
402	構造解析部
403	内容解析部
404	索引ファイル
405	文字列検索装置
3501	処理装置
3502	文書取得部
3503	二次記憶装置
3504	内容推定部
3505	表示用インデックス

3506 検索エンジン
 3507 検索用インデックス
 3508 ビュー生成部

3509 入力装置
 3510 表示装置

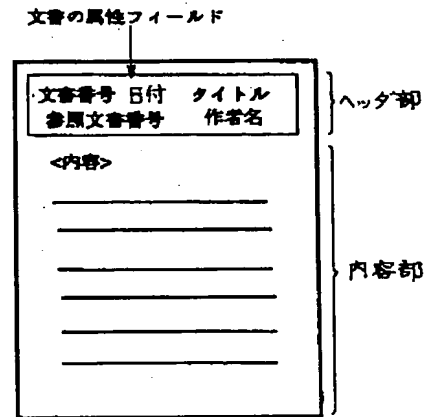
【図1】

参照関係を有する文書群の例を示す図



【図2】

文書のデータ構造の例を示す図



【図3】

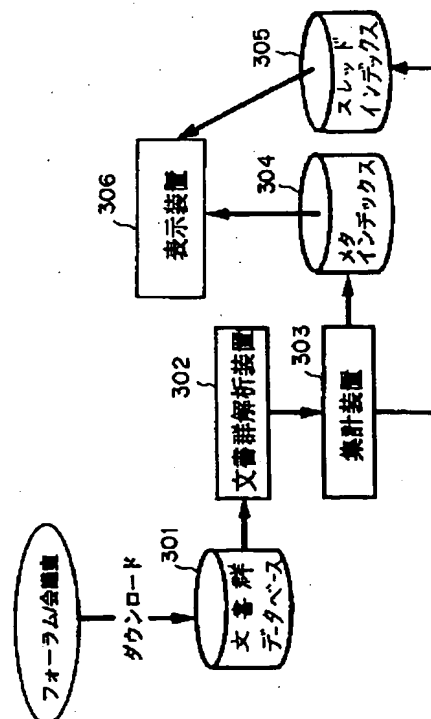
本発明の実施の形態のシステム構成図
 (その1)

【図7】

索引ファイルのデータ構造を示す図

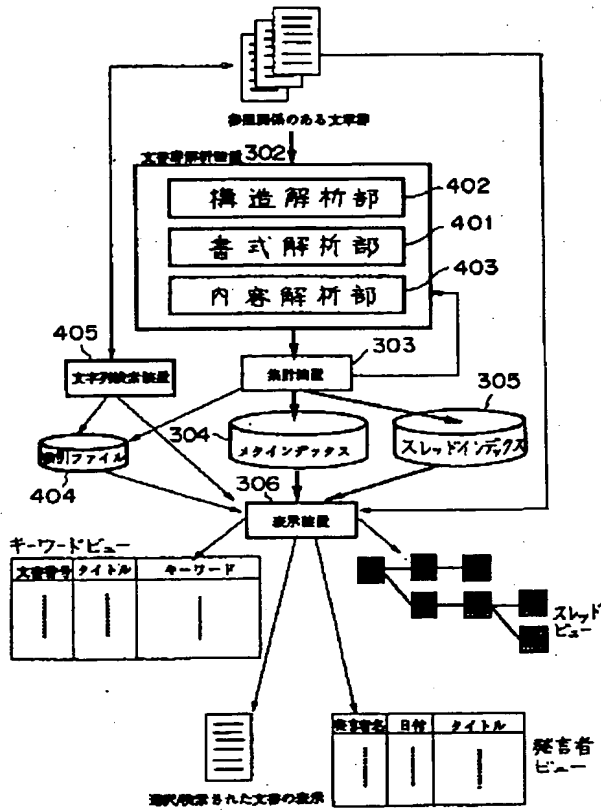
索引欄	索引欄を含むキーワード	スレッド番号*
インターネット	・インターネットスタート ・インターネットへの接続	00496 00218
コントロール	・コントロールパネル	00518
ドメイン	・ドメインネームサーバ ・国のドメイン	00128 00476
—	—	—

*スレッド番号はスレッドのルートの文書番号



【図4】

本発明の実施の形態のシステム構成図
(その2)



【図12】

カラーテーブルの例を示す図

カ-ト 文書番号	色番号	タイトル	文書番号
:	:	:	:
001	#1	(1471ビタ)	001, 002, 004
	#2	(1471ビタ 1)	003
	#3	(1471ビタ 2)	005, 006
:	:	:	:

【図5】

メタインデックスのデータ構造を示す図

文書番号	色番号	タイトル	キーワード	日付	文書番号
001	001	001	001	001	001
002	002	002	002	002	002
003	003	003	003	003	003
004	004	004	004	004	004
005	005	005	005	005	005
006	006	006	006	006	006
007	007	007	007	007	007

【図10】

文書番号のスレッドインデックスへの登録の動作説明図

- エントリの登録
- (1)
 - (001) (2)
 - (001 002) (3)
 - (001 (002)) (4)
 - (001 (002 003)) (5)
 - (001 (002 003) 004) (6)
 - (001 (002 003) (004)) (7)
 - (001 (002 003) (004 005)) (8)
 - (001 (002 003) (004 (005))) (9)
 - (001 (002 003) (004 (005 006))) (10)

【図6】

【図8】

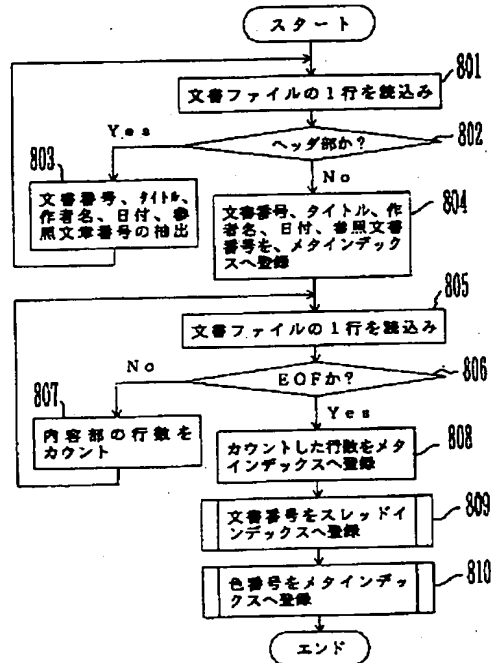
【図36】

スレッドインデックスのデータ構造を示す図書式解析部と構造解析部の動作フローチャート ユーザインデックスの構成図

ルート 文書番号	文書数	スレッドの構造
...		
001	8	(001(002 003)(004(005 006)))
007	5	(007 008 009(011 012))
.....		

001 002-003
004-005-006
007 008 009
011-012

リストの意味: (親文書番号 子文書番号/サブツリー...)



UserID	名前	職務	発言数(回答数)
IX001	津田	津	51 (26)
IX002	内野	内	42 (24)

【図37】

文書インデックスの構成図

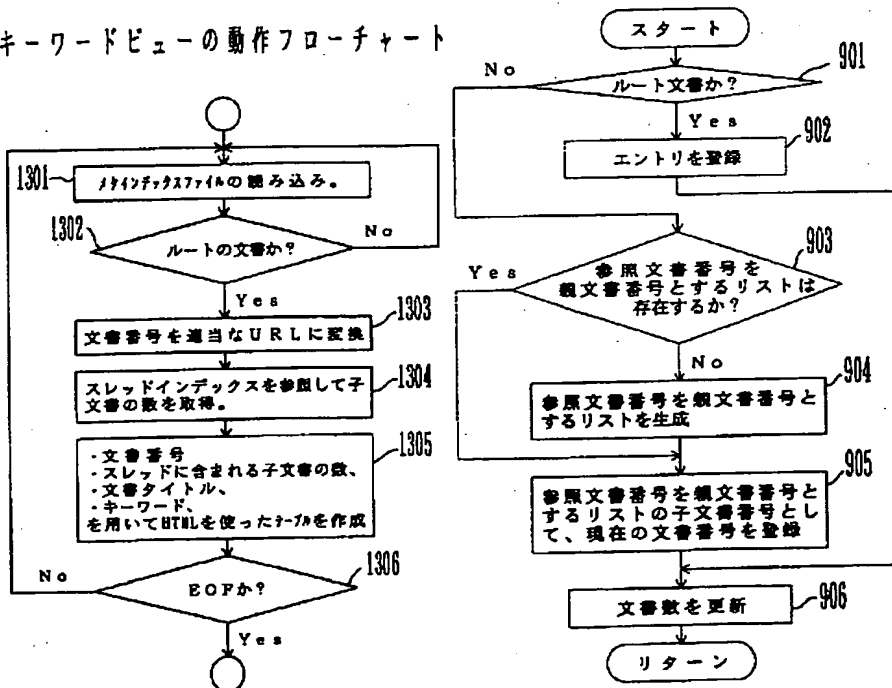
記事番号	Q	A	T
タイトルの識別番号	1	1	2
ルートまでのパス	001	002,001	003,002,001
参照子数	2	1	0
日付	97/9/5	97/9/6	97/9/6
タイトル	プリンタの設定がわかりません!	Re:プリンタの設定がわかりません!	ありがとうございます!
UserID	AAA	BBB	AAA
文書ID	001	002	003

【図9】

文書番号のメタインデックスへの登録の動作フローチャート

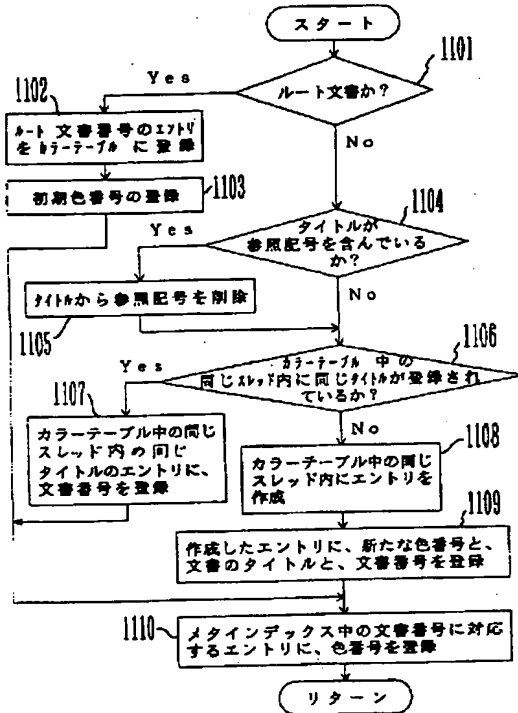
【図13】

キーワードビューの動作フローチャート



【図11】

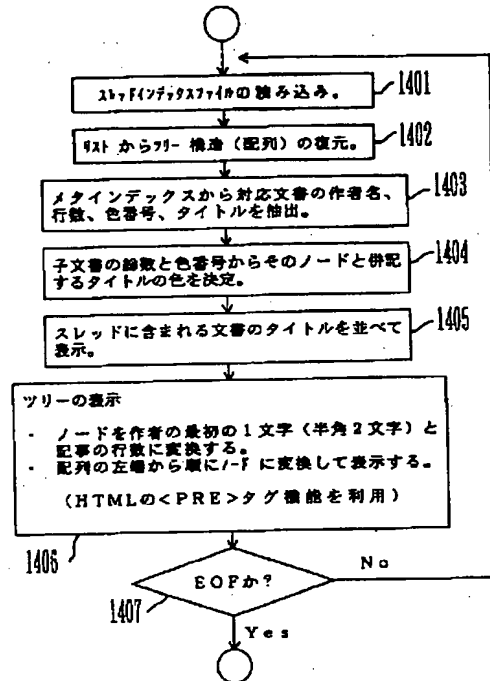
色番号登録の動作フローチャート



【図15】

【図14】

スレッドビューの動作フローチャート



【図16】

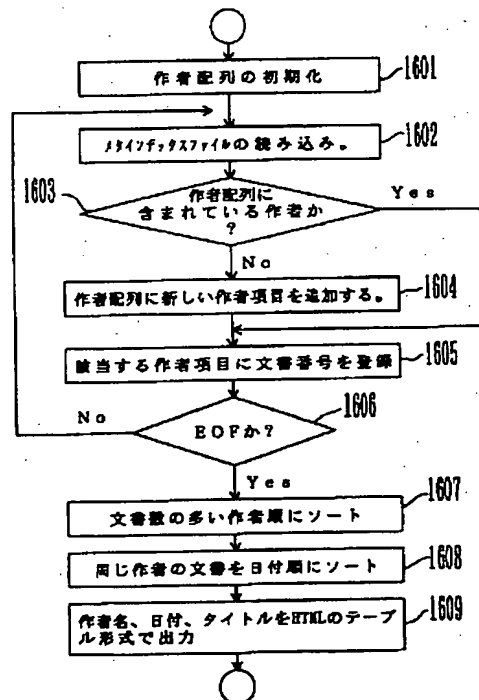
スレッドビューの制御用配列の例を示す図

ツリー構造のリスト：(001(002 003)(004(005 006)))

	1	2	3	4
1	001	002	003	空白
2	空白	004	005	006

ツリー構造を表す配列

発信者ビューの動作フローチャート



【図17】

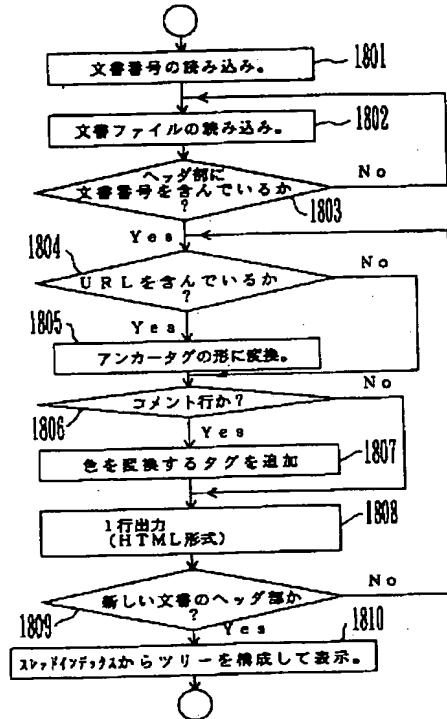
【図18】

【図38】

発言者ビューの制御用配列の例を示す図 発言内容表示の動作フローチャート スレッドインデックスの構成図

作者配列の例

作者 A	001, 005, 007, 008
作者 B	011, 012, 013
作者 C	003, 009, 010
作者 D	004, 006
作者 E	002



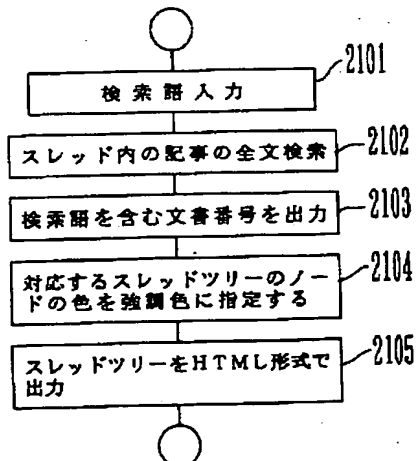
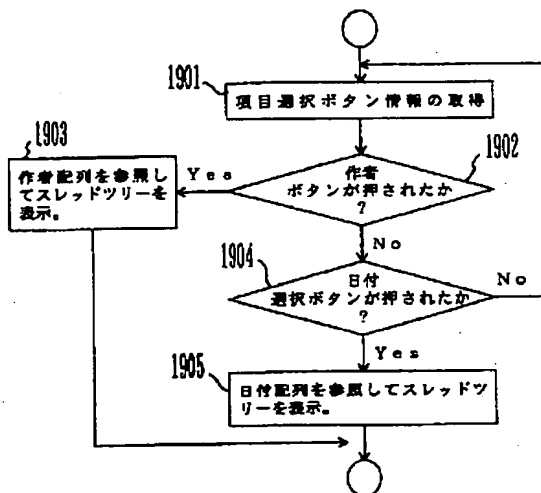
スレッドインデックス

内容リスト	発言者UID	作者数	文書数	スレッドの本数	スレッドFID
(0001, 0002, 0003, 0004)	0001	4	6	(001 (002 003) (004 (005 006)))	001
	0002	4	5	(007 008 009 (011 012))	007

【図19】

【図21】

作者別/日付別色分け表示の動作フローチャート スレッドビューを使った検索結果の強調表示の動作フローチャート



【図20】

作者別 日付別
色分け表示用配列の例を示す図

(b) 日付配列の例

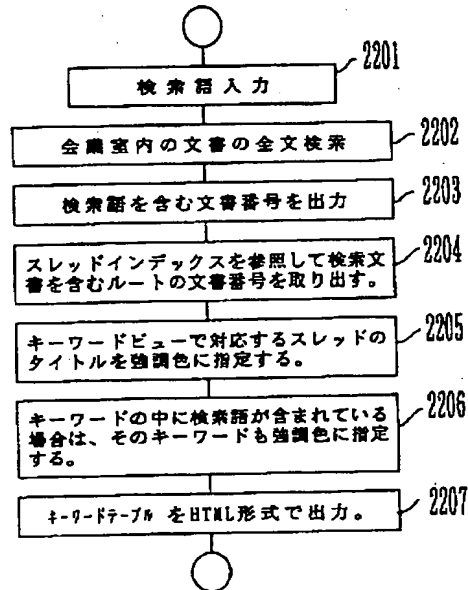
	黄	緑	赤	青	紫
97/04/15	001, 002, 003, 004				
97/04/21		005, 006			
97/04/28			007, 008, 009		
97/05/25				010	
97/05/28					011, 012, 013, 014, 015

(a) 作者配列の例

	黄	緑	赤	青	紫
作者A	001, 005, 007, 008				
作者B		011, 012, 013			
作者C			003, 009, 010		
作者D				004, 006	
作者E					002

【図22】

キーワードビューを使った検索結果の
強調表示の動作フローチャート

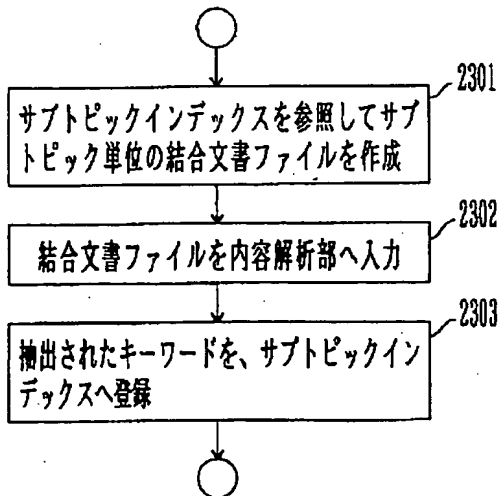


【図24】

サブピックインデックスの例を示す図

【図23】

サブピックからのキーワード抽出の
制御動作フローチャート



キーワード	記事番号	タイトル	色番号	サブ文書番号
...
...	001, 002, 004	(11/11/11)	#1	001
...	003	(11/11/11)	#2	
...	005, 006	(11/11/22)	#3	
...

【図64】

カレンダーインデックスのデータ構成図

日付	文書ID
10/1	001, 002
10/2	003
10/3	004, 005, 006

【図25】

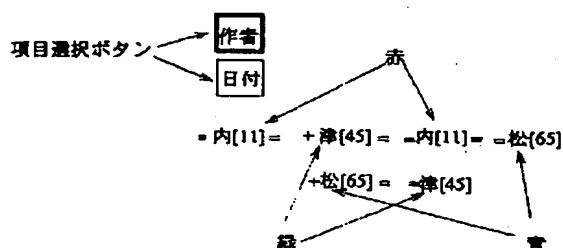
キーワードビューの表示例を示す図

fgal_internet_QAのキーワードビュー		
[スレッドビュー] [キーンビュー]		
ID	タイトル	キーワード
00001	1 はじめに	質問の内容、当会謝辞、発言のタイトル、Internet、師道の 質疑応答、接続形式、用語集(ハードウェア、的確な内容の アドバイス、ばば/総理、Manager
00002	3 最低限のハードウ ェア	最近流行のネットサーフィン、インターネット、デファクト ・スタンダードのネットスケープ・ナビゲーター、ペンティ アムマシン、こっち、ならパソコン、法律の条文、色々なサー ビス、PeopleのWorldTalk、FDDI
00005	3 質問:ホームページ につながらない	プロバイダ、パスワードの確認、使用機種、設定、ネットス ケープのデフォルト、ダイヤルアップ接続、使用ソフト FM*DPSP5、ユーザ名
00008	6 インターネットス タート	プロバイダ、ホームページ、料金、テレホーダイの時間帯 ・WWW・アクセス、ISDN、アナログ288、回線速度、nifty のコマンド
00014	1 皆さん有り難う	リーヤさん推薦のプロバイダ、勝手気便アバウトネットワー カー、明日電話、金の社員書店、まどレス、いろいろチャレ ンジ、せ名古屋、ROROS、この間、子会社
00015	24 ニフティのダイア ルアップ接続	@niftyserve、つげ、スクリプト、SubManager、or、金星人 ・NIFTYServe、インターネット、ウィザード、kazusi
00016	1 プロバイダ選び	勝手気便アバウトネットワーカー、オンラインサイ、リーヤ さん推薦、ホームページの設定、資料請求、シアップ、固定 料金、程度信頼、結構痛、ALLESNET
00030	9 プロバイダー---目 宅PC間の速度	サーバーサーバー自宅PC、@niftyserve、プロキシ、TAK、 ping、MNP5、下の方のレイヤのキャッシュ、SubManager、 金星人、モデムコマンド
00040	24 ニフ接続のメール	インターネットのメール、@niftyserve、SubManager、or ・ばば/総理、金星人、ニフ接続の人、NIFTYServe、サービ スの中、kazusi
Windows95のサポート、Windows用のソフト、まんま、		

【図29】

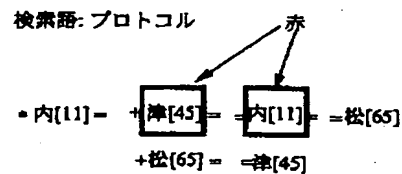
スレッドビューを用いた文書属性

「作者」の強調(色別)表示の例を示す図



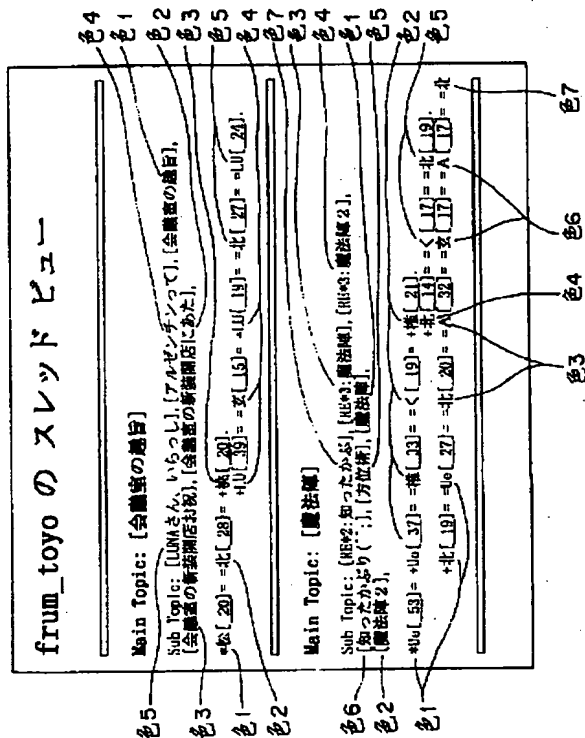
【図31】

スレッドビューを用いた文字列

「プロトコル」を含むノードの強調表示の
例を示す図

【圖 26】

スレッドビューの表示例を示す図



【図 3 2】

キーワードビューを用いた文字列
「プロトコル」を含むスレッドの
強調表示の例を示す図

検索語: プロトコル	
会議室名: FGL インターネット	
タイトル	キーワード
A 先生がケーブルを	B 大学 C 校舎 インターネットのはじまり
プロトコル	connect ftp UNIXマシンのプロトコル
Internetの周縁	プロトコルの種類 tcp ftp

【図 30】

会議室内の全発言の検索(入力)表示例を示す図

fgal_internet_Q&A会議室の文書の検索ページ

このサーバで、fgal_internet_Q&A会議室の文書リソースを検索します。
 次のフォームで、検索文字列および検索条件を入力して [Search] ボタンをクリックして下さい
 検索結果に別し絞り込みもできるので、最初から厳しい条件で検索しない方が良いでしょう。

UNIX

I

I

の ☐ うちどれか ☐ を含む、

ネットスケープ

インターネット

ホームページ

を含まない文書を検索します。

ただし、アルファベットの大小文字は区別 ☐ し、空白系文字列は

無視せず ☐ 検索します。

最初のページに戻る

【图 3 3】

会議室内の全発言の検索(結果出力)
表示例を示す図

Result: FOUND

このサーバの /usr/Internet/WWW/WWW の文書リソースの中で、検索式 (UNIX) の「ネットワーク」および「インターネット」を「ホームページ」を満たすものは14/519個ありました。

1. 00004 : RE : 2:プロトコル
2. 00001 : RE:cbad は...
3. 00204 : RE: ニアのダウンロード法教えて(INTERNET
4. 00212 : RE : ニアのダウンロード法教えて(INTERNE
5. 00263 : RE: ニアを載ったのにいひ(;)
6. 00270 : RE : 2:0
7. 00308 : TIV
8. 00330 : RE-RE : ニアのダウンロード法教えて(INTER
9. 00342 : UNIX用のPLUS
10. 00342 : RE: プロセス管理を防ぐ
11. 00376 : RE: モニターが現れ無いでしょ? (;)
12. 00463 : RE: Q. tarの付く拡張子ってどうすれば
13. 00475 : RE: RE : 2:Q. tarの付く拡張子ってどうすれば
14. 00480 : RE : 4:Q. tarの付く拡張子ってどうすれば

絞り込み検索

この検索結果の中から、さらに条件を与えて絞り込みをすることができます。以下のフォームで条件を指定して「Search」のボタンをクリックして下さい。新たに別の検索を行う場合には「最初のページ」に戻って行って下さい。

【図27】

発言者ビューの表示例を示す図

	9月	08日: [RE:4:SANNYNET] 存在の方おられますか?
MURA	9月	11日: [RE:NIのAdmin] 機能
	30日	[RE:スネケ3.01.06でQuickTime]
	10月	1日: [RE:スネケ3.01.06でQuickTime] [RE:NN3Ja.0のインストール出来
かんちゃん ★	10月	2日: [RE:スネケ3.01.06でQuickTime] [RE:スネケ3.01.06でQuickTime]
	7月	25日: [RE:0.05でnameを消さない]
	27日	[RE:0.05でnameを消さない]
	9月	29日: [RE:強子 rackについて] [RE:匿名メール]
	30日	[RE:強子 rackについて]
よしちや	10月	1日: [RE:強子 rackについて]
	5月	08日: [RE:スネケ3.01.06のダイヤルアップ機能]
	18日	[RE:スネケ3.01.06のダイヤルアップ機能]
	6月	05日: [RE:4:SANNYNET] 存在の方おられますか?
荒木 修	6月	06日: [RE:4:SANNYNET] 存在の方おられますか?
	08日	[RE:ActiveXの初歩的質問です]
	08日	[RE:ActiveXの初歩的質問です]
	09日	[RE:ActiveXの初歩的質問です]
酒匂 義弘	9月	02日: [RE:ActiveXの初歩的質問です]
	02日	[RE:ActiveXの初歩的質問です]
	03日	[RE:ActiveXの初歩的質問です]
	03日	[RE:ActiveXの初歩的質問です]
	03日	[RE:ActiveXの初歩的質問です]
Mue	8月	14日: [RE:ActiveXの初歩的質問です]
	14日	[RE:ActiveXの初歩的質問です]
	15日	[RE:ActiveXの初歩的質問です]
	16日	[RE:ActiveXの初歩的質問です]
安藤 貞之	7月	27日: [RE:ActiveXの初歩的質問です]
	28日	[RE:ActiveXの初歩的質問です]
	28日	[RE:ActiveXの初歩的質問です]
	30日	[RE:ActiveXの初歩的質問です]
N-DITE	8月	18日: [RE:ActiveXの初歩的質問です]
	19日	[RE:ActiveXの初歩的質問です]
	20日	[RE:ActiveXの初歩的質問です]

【図34】

サブトピックから抽出した
キーワードの表示例を示す図

メイントピック: 何から質問すれば

サブトピック	キーワード
A 先生がケーブルを	B 大学 C 校舎 インターネットのはじまり
プロトコル	ethet Rtp UNIXマシンのプロトコル

【図39】

QA インデックスの構成図

QA_ID	Question	Answer	Thanks	Max AnswerUID
:	:	:	:	:
Q0011	001	002	003	KX001
:	:	:	:	:

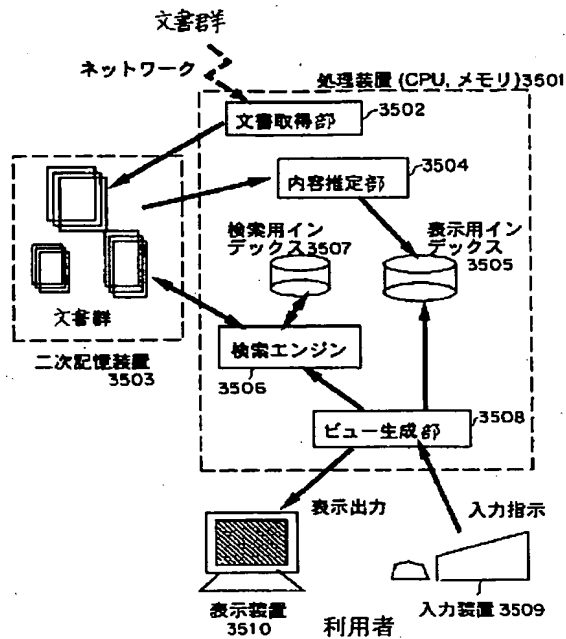
* QA_IDは文字列QAとスレッドIDと
通し番号からなる

* Question文書からThanks文書までの
パスにある文書を全てAnswer文書として
登録する

* Answer文書群の中で一番多いUIDを
MaxAnswerUIDとして登録する

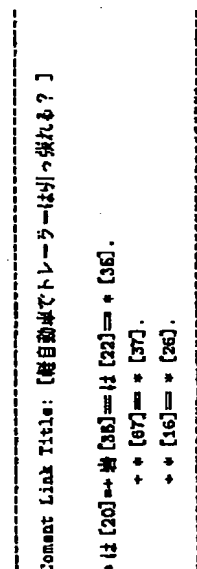
【図35】

本発明の他の実施の形態(第2の実施の形態)
の構成図



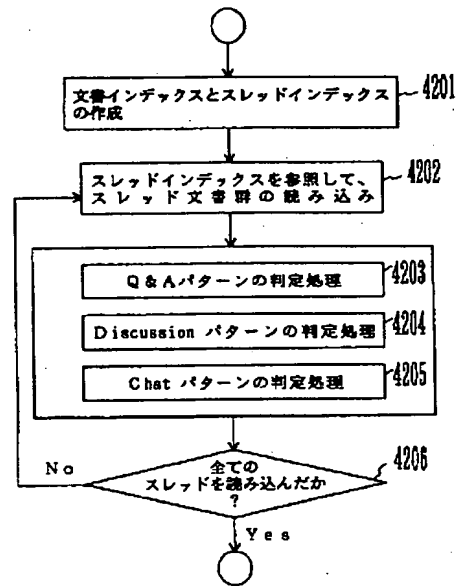
【図46】

Q and A パターンの判定処理が推測
するスレッド構造の例を示す図



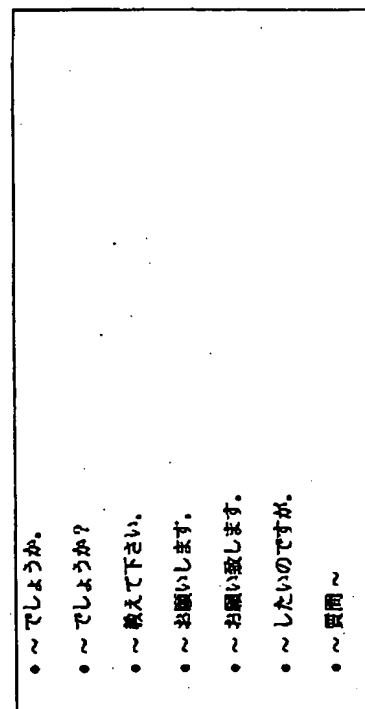
【図42】

内容推定部の動作フローチャート



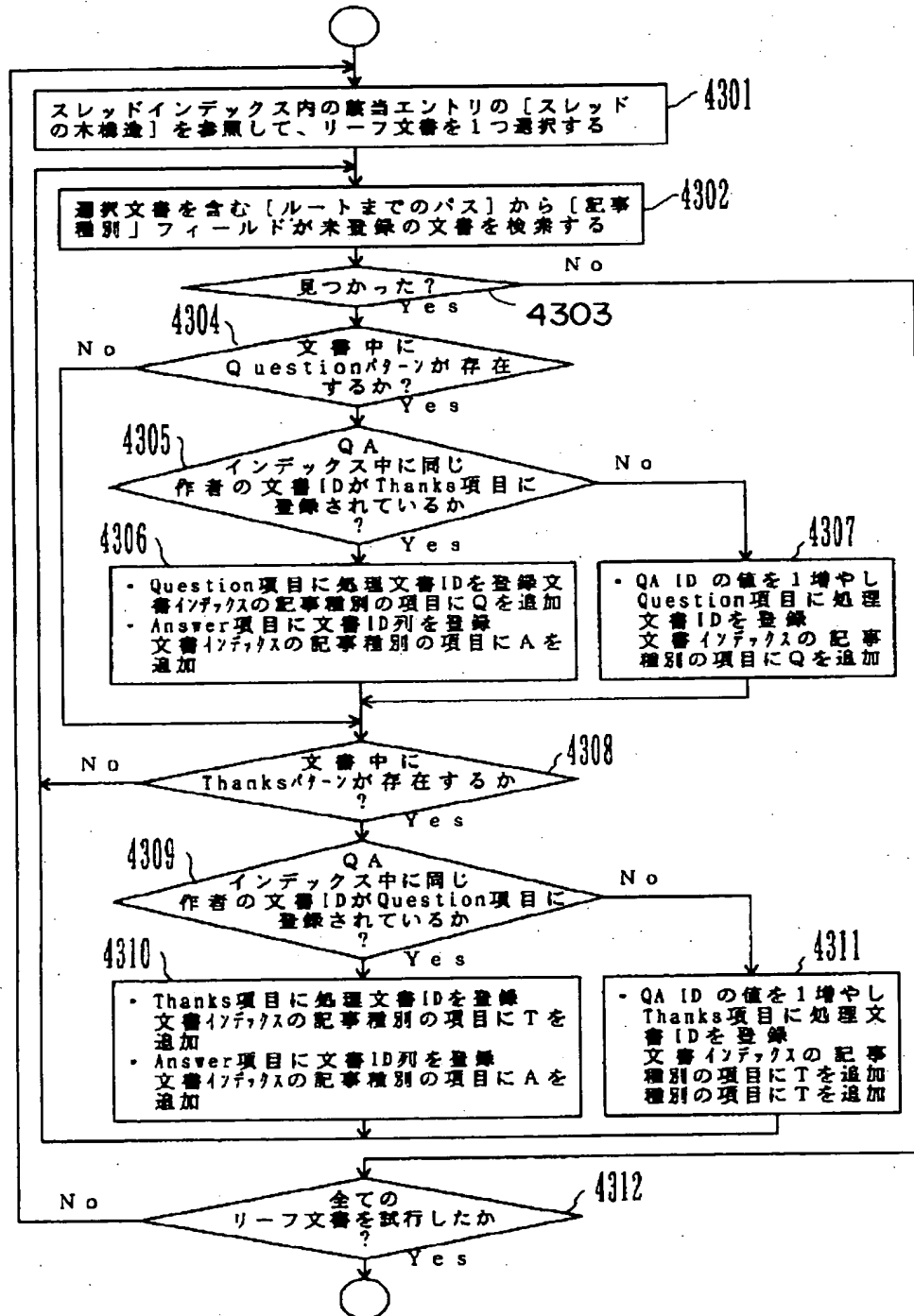
【図44】

Q (質問) 文書に含まれるパターンの例を示す図



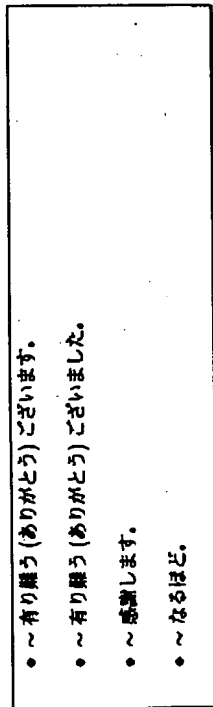
【図43】

Q and A パターン判定処理の動作フローチャート



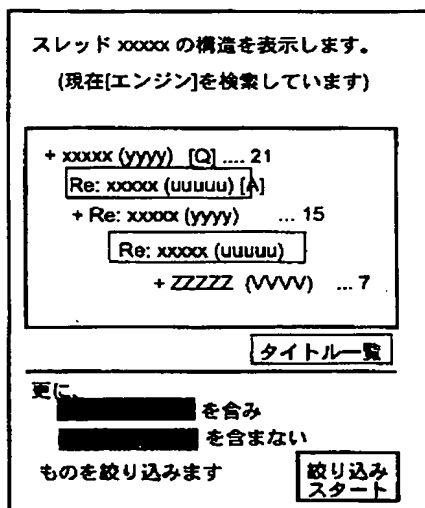
【図45】

お礼文書に含まれるパターンの例を示す図

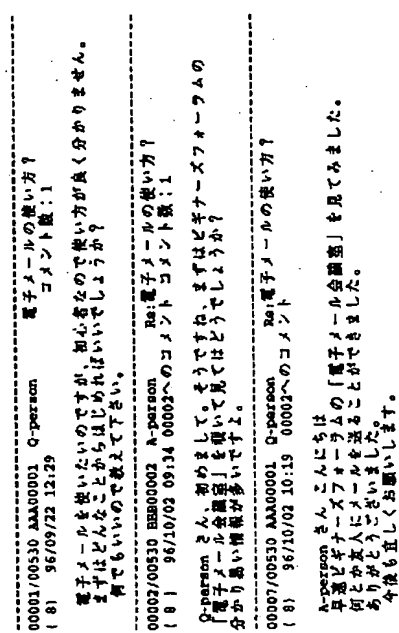


【図61】

出力結果2:スレッド構造表示を示す図

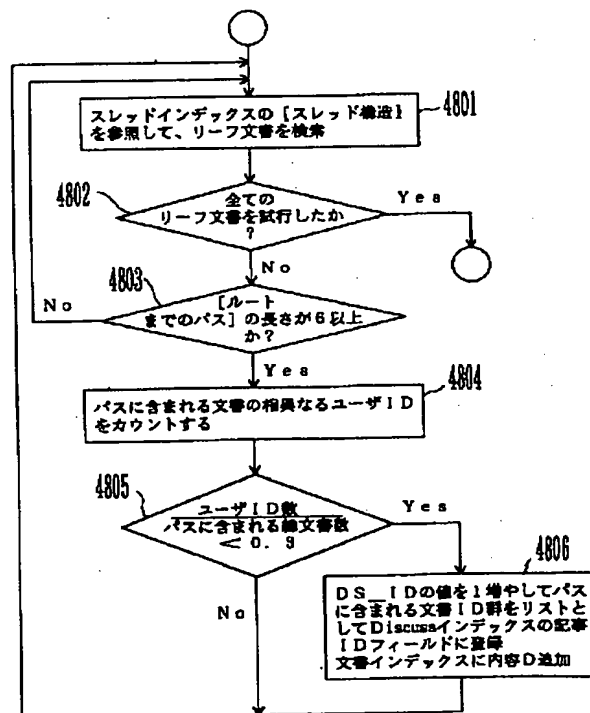


【図47】

Q and A ハターの判定処理が
推測するスレッド文書群の例を示す図

【図48】

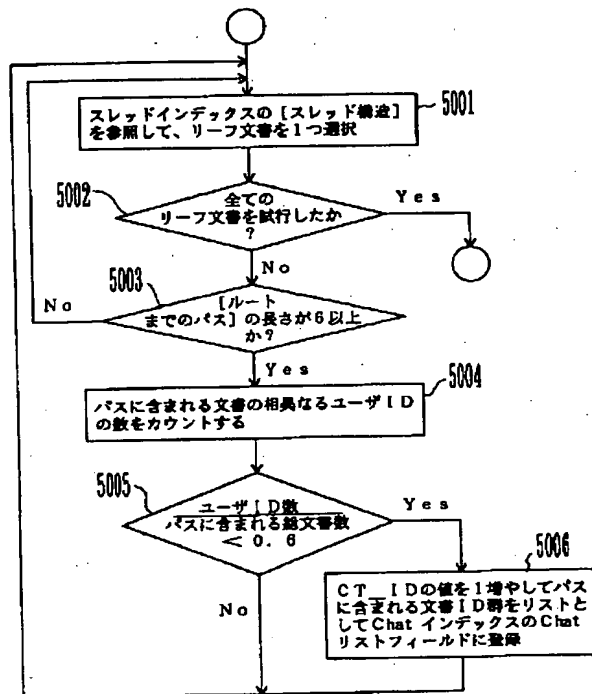
論争パターンの特定処理の動作フローチャート



【图 50】

議論パターンの判定処理の動作フローチャート

スレッド構造の例を示す図



【圖 5 2】

オリジナルのスレッド構造の表示例を示す図

スレッド構造の例を示す図

03128: [45: 小川和芳] お久しぶりのA Z - 1
03129: [9: 小川和芳]
03136: [29: G a l a]
03153: [10: パーティバル]
03167: [29: G a l a]
03181: [17: パーティバル]
03182: [24: 小川和芳]
03154: [10: パーティバル]
03184: [10: 小川和芳]
03189: [21: EF-3CIVIC]
03201: [23: パーティバル]
03209: [22: EF-3CIVIC]
03210: [17: 小川和芳]
03200: [19: パーティバル]
03137: [44: やまやま]
03155: [20: パーティバル]
03183: [27: 小川和芳]

【图 5 3】

オリジナルのスレッド構造の表示例を
示す図

$$\begin{aligned} * \text{小} [45] &= + \text{小} [9] = + G [29] = + \text{小} [10] = G [28] = + \text{小} [17], \\ &+ \text{小} [24]. \\ &+ \text{小} [10] = + \text{小} [10] = + EF [21] = + \text{小} [23] = + EF [32], \\ + \text{小} [44] &= + \text{小} [20], \quad + \text{小} [19], \quad + \text{小} [17], \\ &+ \text{小} [27]. \end{aligned}$$

【图 5 4】

参照ノード数により縮退したスレッド構造の
表示例を示す図

+ 03128: [46: 小川和芳] お久しぶりのAZ-1 (17)
+ 03129: [9: 小川和芳] Re: お久しぶりのAZ-1 (13)
+ 03136: [29: G a l a] おっ、久々にAZ-1の文字が (5)
03154: [10: パーツィバル] Re: お久しぶりのAZ-1
03184: [10: 小川和芳] 5、5 Jのアルミ
+ 03189: [21: EF-3CIVIC] Re: 5、5 Jのアルミ (4)

【圖 5 6】

時間区間で取り出したスレッド構造の表示例を
示す図

【図 5 5】

表示例を示す図

10/1 10/2 10/3 10/4 10/5

*小 ———— 小, G ———— B, G ———— B
 + ———— 小
 + B, 小 ———— EF, B ———— EF

【图 6 6】

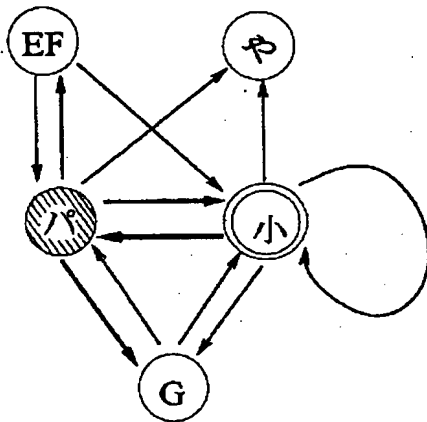
発言者配列のデータ構成図

+ 03128: [45: 小川和芳] お久しぶりのAZ-1
 + 03136: [29: G a i a] おっ、久々にAZ-1の文字が
 + 03184: [10: 小川和芳] 5、5Jのアルミ
 03210: [17: 小川和芳] 手が掛かりすぎ?
 03183: [27: 小川和芳] スプリングとインタークーラー

額発言 子発言	AAA	BBB	CCC
AAA	0	5	2
BBB	2	1	1
CCC	0	1	0

【図57】

作者をノードとしたグラフ構造を示す図



【図60】

出力結果1: スレッド一覧を示す図

車に関する会議室から[エンジン]を検索した結果です。
全部で314記事ありました。
上位10スレッドを表示します。

タイトル	作者	日付	サイズ	内容
xxxxxx	yyy	10/3	22 (14KB)	QA, DC
.....				

→ 次の10スレッド

更に、
☐ を含み
☐ を含まない
 ものを絞り込みます

【図58】

ネットワークニュース検索システムの表示例を示す図

XXニュースサービスの、
車に関する会議室から、次の
条件を満たす記事を検索します。

☐ を含み
☐ を含まない

対象記事の種別
☒ 全記事 ☐ Q&A

対象記事の日付
☒ 全て ☐ 一月以内 ☐ 一週以内

最終更新 97.10.10 3:10AM
全12363記事

【図62】

出力結果3: 時間区間スレッド表示を示す図

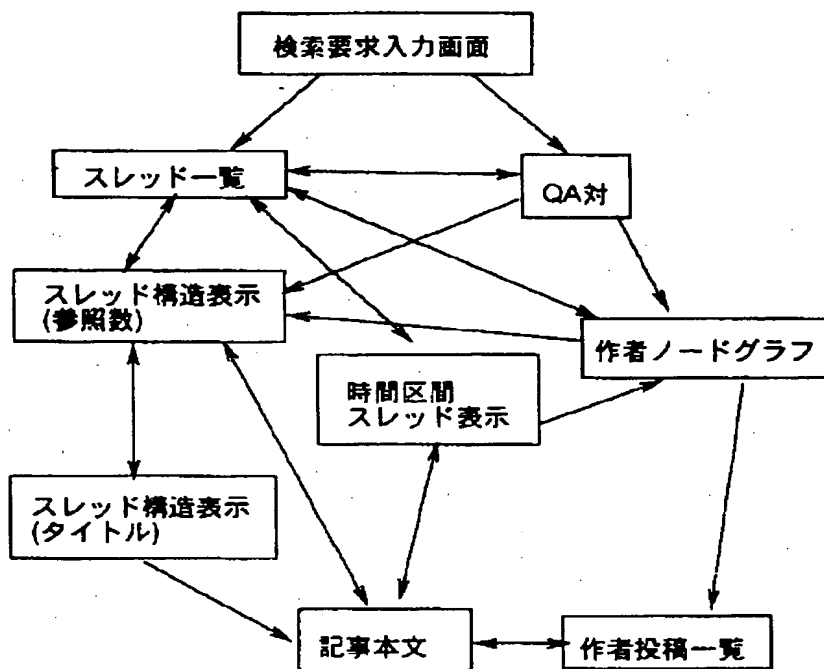
時間区間スレッド表示

10/1 Sun	2 Mon	3 Tue	4 Wed	5 Thu	6 Fri	7 Sat
uu vv	xx	yy		uu	vv	

スケジュール						
日	月	火	水	木	金	土
	15:00 会議	新機 発表	11:30 グループ 会議		出張	

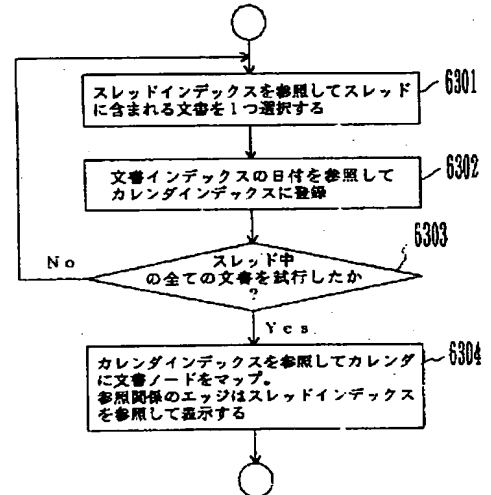
【図59】

検索画面の一覧を示す図



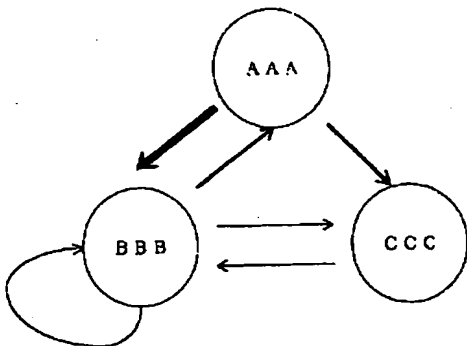
【図63】

時間区間スレッド表示の作成フローを示す図



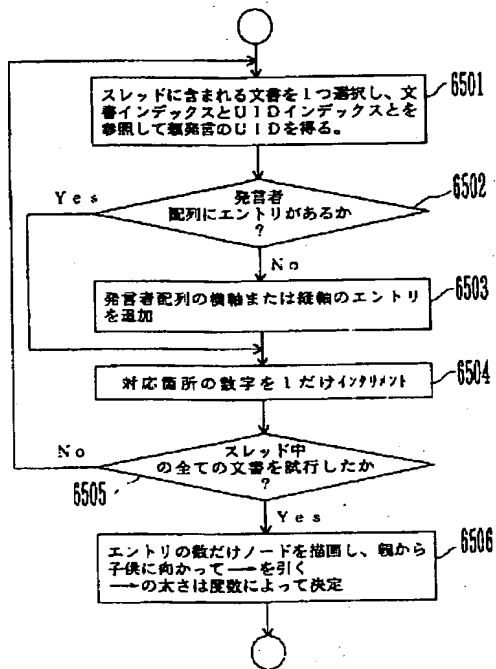
【図67】

作者ノードグラフの作成説明図

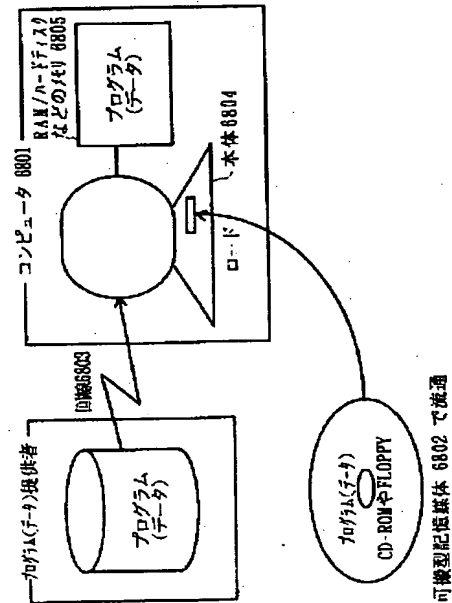


【図65】

作者ノードグラフの作成フローを示す図



【図68】

本実施の形態を実現するプログラムが
記録された記録媒体の説明図

フロントページの続き

(51) Int. Cl. 6

識別記号

FI
G 0 6 F 15/403

3 8 0 D